

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification:

A61K

A2

(11) International Publication Number:

WO 99/42076

(43) International Publication Date:

26 August 1999 (26.08.99)

(21) International Application Number:

PCT/US99/03268

(22) International Filing Date:

17 February 1999 (17.02.99)

(30) Priority Data:

09/025,197 18 February 1998 (18.02.98) US

09/072,967 5 May 1998 (05.05.98) US

(71) Applicant: CORINA CORPORATION [US/US], 1124 Columbia Street, Seattle, WA 98104 (US).

(72) Inventors: REED, Steven, G.; 2845 122nd Place, N.E., Bellevue, WA 98005 (US); SKEIKY, Yasin, A.; W.; 8327 25th Street, Seattle, WA 98107 (US); DILLON, David, C.; 21607 N.E. 24th Street, Redmond, WA 98053 (US); CAMPOS-NEIRO, Antonio; 9308 Midship Court, N.E., Bainbridge Island, WA 98110 (US); HOUGHTON, Raymond; 2636 242nd Place, S.E., Bothell, WA 98021 (US); VEDVICK, Thomas, S.; 124 South 300th Place, Federal Way, WA 98003 (US); TWARDZIK, Daniel, R.; 10195 South Beach Drive, Bainbridge Island, WA 98110 (US); LODES, Michael, J.; 9223 36th Avenue S.W., Seattle, WA 98126 (US); HENDRICKSON, Ronald, C.; 4114 S.W. Charlestown Street, Seattle, WA 98116 (US).

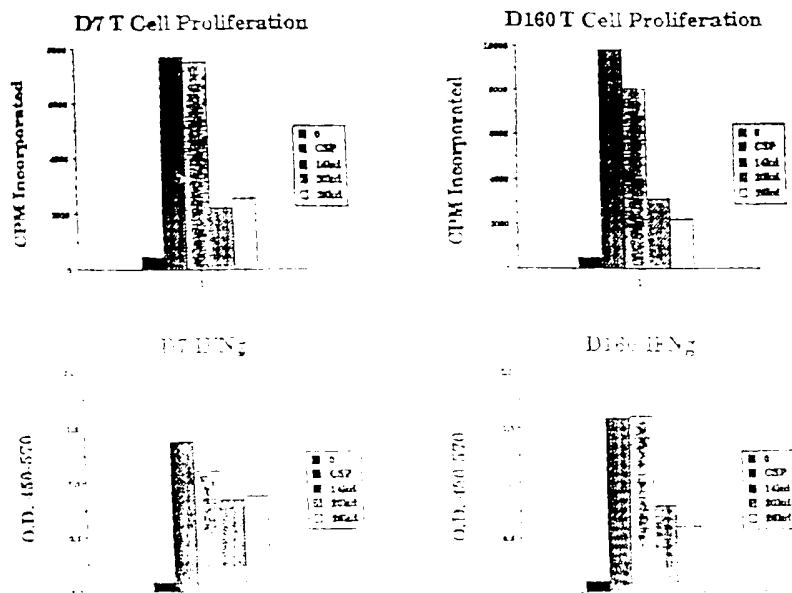
(74) Agents: CORUZZI, Laura, A. et al., Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IL, IT, MC, NL, PT, SE), OAPI patent (BF, BI, CE, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report

(54) Title: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PC L on the front pages of pamphlets publishing international applications under the PC L.

AI	Angola	IS	Israel	IS	Isle of Man	SI	Slovenia
AM	Armenia	IT	Italy	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AX	Åland Islands	GA	Gabon	LV	Latvia	SZ	Switzerland
AZ	Azerbaijan	GB	Great Britain	MG	Madagascar	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Malawi	TL	Timor-Leste
BD	Bangladesh	GN	Guinea	MK	Macedonia	TM	Turkmenistan
BE	Belgium	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HR	Croatia	MT	Malta	TT	Trinidad and Tobago
BI	Burundi	IE	Ireland	MS	Montserrat	UA	Ukraine
BJ	Benin	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NI	Nicaragua	VN	Vietnam
CG	Congo	KE	Kenya	NE	Niger	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NG	Nigeria	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			P	Panama		
				F	Finland		

COMPOUNDS AND METHODS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS

5 CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of U.S. Application No. 9/025,197, filed February 18, 1998; which is a continuation-in-part of U.S. Application No. 08/942,578, filed October 1, 1997; which is a continuation-in-part of U.S. Application No. 08/818,112, filed March 13, 1997; which is a continuation-in-part of U.S. Application No. 08/730,510, filed October 11, 1996; which claims priority from PCT Application No. PCT/US 96/14674, filed August 30, 1996; and is a continuation-in-part of U.S. Application No. 08/680,574, filed July 12, 1996; which is a continuation-in-part of U.S. Application No. 08/659,683, filed June 5, 1996; which is a continuation-in-part of U.S. Application No. 08/620,874, filed March 22, 1996, now abandoned; which is a continuation-in-part of U.S. Application No. 08/533,634, filed September 22, 1995, now abandoned; which is a continuation-in-part of U.S. Application No. 08/523,436, filed September 1, 1995, now abandoned.

TECHNICAL FIELD

20 The present invention relates generally to detecting, treating and preventing *Mycobacterium tuberculosis* infection. The invention is more particularly related to polypeptides comprising a *Mycobacterium tuberculosis* antigen, or a portion or other variant thereof, and the use of such polypeptide for diagnosing and vaccinating against *Mycobacterium tuberculosis* infection.

25 BACKGROUND OF THE INVENTION

Tuberculosis is a chronic infectious disease that is generally caused by infection with *Mycobacterium tuberculosis*. It is a major disease in developing countries as well as in industrialized countries. Tuberculosis is caused by

manifested as an acute inflammation of the lungs, resulting in fever and a nonproductive cough. If left untreated, serious complications and death typically result.

Although tuberculosis can generally be controlled using extended antibiotic therapy, such treatment is not sufficient to prevent the spread of the disease. Infected individuals may be asymptomatic, but contagious, for some time. In addition, although compliance with the treatment regimen is critical, patient behavior is difficult to monitor. Some patients do not complete the course of treatment, which can lead to ineffective treatment and the development of drug resistance.

Inhibiting the spread of tuberculosis requires effective vaccination and accurate, early diagnosis of the disease. Currently, vaccination with live bacteria is the most efficient method for inducing protective immunity. The most common *Mycobacterium* employed for this purpose is *Bacillus Calmette-Guerin* (BCG), an avirulent strain of *Mycobacterium bovis*. However, the safety and efficacy of BCG is a source of controversy and some countries, such as the United States, do not vaccinate the general public. Diagnosis is commonly achieved using a skin test, which involves intradermal exposure to tuberculin PPD (protein-purified derivative). Antigen-specific T cell responses result in measurable induration at the injection site by 48-72 hours after injection, which indicates exposure to *Mycobacterial* antigens. Sensitivity and specificity have, however, been a problem with this test, and individuals vaccinated with BCG cannot be distinguished from infected individuals.

While macrophages have been shown to act as the principal effectors of *M. tuberculosis* immunity, T cells are the predominant inducers of such immunity. The essential role of T cells in protection against *M. tuberculosis* infection is illustrated by the frequent occurrence of *M. tuberculosis* in AIDS patients, due to the depletion of CD4⁺ T cells associated with human immunodeficiency virus (HIV) infection. *Mycobacterium* reactive CD4⁺ T cells have been shown to be potent producers of gamma-interferon (IFN- γ), which, in turn, has been shown to trigger the antimycobacterial effects of macrophages in mice. While the role of IFN- γ in humans is

to inhibit *M. tuberculosis* infection. Furthermore, it is known that IFN- γ stimulates human macrophages to make 1,25-dihydroxy-vitamin D₃. Similarly, IL-12 has been shown to play a role in stimulating resistance to *M. tuberculosis* infection. For a review of the immunology of *M. tuberculosis* infection see Chan and Kaufmann in
 5 *Tuberculosis: Pathogenesis, Protection and Control*, Bloom (ed.), ASM Press, Washington, DC, 1994.

Accordingly, there is a need in the art for improved vaccines and methods for preventing, treating and detecting tuberculosis. The present invention fulfills these needs and further provides other related advantages.

SUMMARY OF THE INVENTION

Briefly stated, this invention provides compounds and methods for preventing and diagnosing tuberculosis. In one aspect, polypeptides are provided comprising an immunogenic portion of a soluble *M. tuberculosis* antigen, or a variant of
 15 such an antigen that differs only in conservative substitutions and/or modifications. In one embodiment of this aspect, the soluble antigen has one of the following N-terminal sequences:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-
 Gln-Val-Val-Ala-Ala-Leu (SEQ ID No. 120)
- (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-
 Ser (SEQ ID No. 121)
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Ile-Glu-Ala-
 Ala-Lys-Glu-Gly-Arg (SEQ ID No. 122)
- (d) Tyr-Lys-Tyr-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Ile-Gly-
 Pro (SEQ ID No. 123)
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val;
 (SEQ ID No. 124)
- (f) Ala-Glu-Gln-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID

- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly; (SEQ ID No. 127)
- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn; (SEQ ID No. 128)
- (j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID No. 134)
- (k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID No. 135) or
- (l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 136)

wherein Xaa may be any amino acid.

In a related aspect, polypeptides are provided comprising an immunogenic portion of an *M. tuberculosis* antigen, or a variant of such an antigen that differs only in conservative substitutions and/or modifications, the antigen having one of the following N-terminal sequences:

- (m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID No. 137) or
- (n) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID No. 129)

wherein Xaa may be any amino acid.

In another embodiment, the soluble *M. tuberculosis* antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID Nos. 1, 2, 4-10, 13-25, 52, 99 and 101, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos. 1, 2, 4-10, 13-25, 52, 99 and 101 or a complement thereof under moderately stringent conditions.

In a related aspect, the polypeptides comprise an immunogenic portion of

amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID Nos.: 26-51, 138, 139, 163-183, 201, 240, 242-247, 253-256, 295-298, 309, 316, 318-320, 322, 324, 328, 339, 333, 335, 337, 339 and 341, the complements of said sequences, and DNA sequences that hybridize to a sequence
5 recited in SEQ ID Nos.: 26-51, 138, 139, 163-183, 201, 240, 242-247, 253-256, 295-298, 309, 316, 318-320, 322, 324, 328, 329, 333, 335, 337, 339 and 341 or a complement thereof under moderately stringent conditions.

In related aspects, DNA sequences encoding the above polypeptides, expression vectors comprising these DNA sequences and host cells transformed or
10 transected with such expression vectors are also provided.

In another aspect, the present invention provides fusion proteins comprising a first and a second inventive polypeptide or, alternatively, an inventive polypeptide and a known *M. tuberculosis* antigen.

Within other aspects, the present invention provides pharmaceutical
15 compositions that comprise one or more of the above polypeptides, or a DNA molecule encoding such polypeptides, and a physiologically acceptable carrier. The invention also provides vaccines comprising one or more of the polypeptides as described above and a non-specific immune response enhancer, together with vaccines comprising one or more DNA sequences encoding such polypeptides and a non-specific immune
20 response enhancer.

In yet another aspect, methods are provided for inducing protective immunity in a patient, comprising administering to a patient an effective amount of one or more of the above polypeptides.

In further aspects of this invention, methods and diagnostic kits are
25 provided for detecting tuberculosis in a patient. The methods comprise contacting dermal cells of a patient with one or more of the above polypeptides and detecting an immune response on the patient's skin. The diagnostic kits comprise one or more of the above polypeptides in combination with an apparatus sufficient to contact the

In yet other aspects, methods are provided for detecting tuberculosis in a patient, such methods comprising contacting dermal cells of a patient with one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347; and detecting an immune response on the patient's skin. Diagnostic kits for use in such methods are also provided.

These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE IDENTIFIERS

Figure 1A and B illustrate the stimulation of proliferation and interferon- γ production in T cells derived from a first and a second *M. tuberculosis*-immune donor, respectively, by the 14 Kd, 20 Kd and 26 Kd antigens described in Example 1.

Figure 2 illustrates the stimulation of proliferation and interferon- γ production in T cells derived from an *M. tuberculosis*-immune individual by the two representative polypeptides TbRa3 and TbRa9.

Figures 3A-D illustrate the reactivity of antisera raised against secretory *M. tuberculosis* proteins, the known *M. tuberculosis* antigen 85b and the inventive antigens Tb38-1 and TbH-9, respectively, with *M. tuberculosis* lysate (lane 2), *M. tuberculosis* secretory proteins (lane 3), recombinant Tb38-1 (lane 4), recombinant TbH-9 (lane 5) and recombinant 85b (lane 6).

Figure 4B illustrates the stimulation of interferon- γ production in a TbH-9-specific T cell clone by secretory *M. tuberculosis* proteins, PPD and recombinant TbH-9.

Figures 5A and B illustrate the stimulation of proliferation and interferon- γ production in TbH9-specific T cells by the fusion protein TbH9-Tb38-1.

Figures 6A and B illustrate the stimulation of proliferation and interferon- γ production in Tb38-1-specific T cells by the fusion protein TbH9-Tb38-1.

Figures 7A and B illustrate the stimulation of proliferation and interferon- γ production in T cells previously shown to respond to both TbH-9 and Tb38-1 by the fusion protein TbH9-Tb38-1.

Figures 8A and B illustrate the stimulation of proliferation and interferon- γ production in T cells derived from a first *M. tuberculosis*-immune individual by the representative polypeptides XP-1, RDIF6, RDIF8, RDIF10 and RDIF11.

Figures 9A and B illustrate the stimulation of proliferation and interferon- γ production in T cells derived from a second *M. tuberculosis*-immune individual by the representative polypeptides XP-1, RDIF6, RDIF8, RDIF10 and RDIF11.

SEQ. ID NO. 1 is the DNA sequence of TbRa1.

SEQ. ID NO. 2 is the DNA sequence of TbRa10.

SEQ. ID NO. 3 is the DNA sequence of TbRa11.

SEQ. ID NO. 4 is the DNA sequence of TbRa12.

SEQ. ID NO. 5 is the DNA sequence of TbRa13.

SEQ. ID NO. 6 is the DNA sequence of TbRa16.

SEQ. ID NO. 7 is the DNA sequence of TbRa17.

SEQ. ID NO. 8 is the DNA sequence of TbRa18.

SEQ. ID NO. 9 is the DNA sequence of TbRa19.

SEQ. ID NO. 10 is the DNA sequence of TbRa24.

SEQ. ID NO. 13 is the DNA sequence of TbRa29.
SEQ. ID NO. 14 is the DNA sequence of TbRa2A.
SEQ. ID NO. 15 is the DNA sequence of TbRa3.
SEQ. ID NO. 16 is the DNA sequence of TbRa32.
5 SEQ. ID NO. 17 is the DNA sequence of TbRa35.
SEQ. ID NO. 18 is the DNA sequence of TbRa36.
SEQ. ID NO. 19 is the DNA sequence of TbRa4.
SEQ. ID NO. 20 is the DNA sequence of TbRa9.
SEQ. ID NO. 21 is the DNA sequence of TbRaB.
10 SEQ. ID NO. 22 is the DNA sequence of TbRaC.
SEQ. ID NO. 23 is the DNA sequence of TbRaD.
SEQ. ID NO. 24 is the DNA sequence of YWCPG.
SEQ. ID NO. 25 is the DNA sequence of AAMK.
SEQ. ID NO. 26 is the DNA sequence of TbL-23.
15 SEQ. ID NO. 27 is the DNA sequence of TbL-24.
SEQ. ID NO. 28 is the DNA sequence of TbL-25.
SEQ. ID NO. 29 is the DNA sequence of TbL-28.
SEQ. ID NO. 30 is the DNA sequence of TbL-29.
SEQ. ID NO. 31 is the DNA sequence of TbH-5.
20 SEQ. ID NO. 32 is the DNA sequence of TbH-8.
SEQ. ID NO. 33 is the DNA sequence of TbH-9.
SEQ. ID NO. 34 is the DNA sequence of TbM-1.
SEQ. ID NO. 35 is the DNA sequence of TbM-3.
SEQ. ID NO. 36 is the DNA sequence of TbM-6.
25 SEQ. ID NO. 37 is the DNA sequence of TbM-7.
SEQ. ID NO. 38 is the DNA sequence of TbM-9.
SEQ. ID NO. 39 is the DNA sequence of TbM-12.
SEQ. ID NO. 40 is the DNA sequence of TbM-13.

SEQ. ID NO. 43 is the DNA sequence of TbH-4.
SEQ. ID NO. 44 is the DNA sequence of TbH-4-FWD.
SEQ. ID NO. 45 is the DNA sequence of TbH-12.
SEQ. ID NO. 46 is the DNA sequence of Tb38-1.
5 SEQ. ID NO. 47 is the DNA sequence of Tb38-4.
SEQ. ID NO. 48 is the DNA sequence of TbL-17.
SEQ. ID NO. 49 is the DNA sequence of TbL-20.
SEQ. ID NO. 50 is the DNA sequence of TbL-21.
SEQ. ID NO. 51 is the DNA sequence of TbH-16.
10 SEQ. ID NO. 52 is the DNA sequence of DPEP.
SEQ. ID NO. 53 is the deduced amino acid sequence of DPEP.
SEQ. ID NO. 54 is the protein sequence of DPV N-terminal Antigen.
SEQ. ID NO. 55 is the protein sequence of AVGS N-terminal Antigen.
SEQ. ID NO. 56 is the protein sequence of AAMK N-terminal Antigen.
15 SEQ. ID NO. 57 is the protein sequence of YYWC N-terminal Antigen.
SEQ. ID NO. 58 is the protein sequence of DIGS N-terminal Antigen.
SEQ. ID NO. 59 is the protein sequence of AEES N-terminal Antigen.
SEQ. ID NO. 60 is the protein sequence of DPEP N-terminal Antigen.
SEQ. ID NO. 61 is the protein sequence of APK7 N-terminal Antigen.
20 SEQ. ID NO. 62 is the protein sequence of DPAS N-terminal Antigen.
SEQ. ID NO. 63 is the deduced amino acid sequence of TbRa1.
SEQ. ID NO. 64 is the deduced amino acid sequence of TbRa10.
SEQ. ID NO. 65 is the deduced amino acid sequence of TbRa11.
SEQ. ID NO. 66 is the deduced amino acid sequence of TbRa12.
25 SEQ. ID NO. 67 is the deduced amino acid sequence of TbRa13.
SEQ. ID NO. 68 is the deduced amino acid sequence of TbRa16.
SEQ. ID NO. 69 is the deduced amino acid sequence of TbRa17.
SEQ. ID NO. 70 is the deduced amino acid sequence of TbRa18.

SEQ. ID NO. 73 is the deduced amino acid sequence of TbRa26.
SEQ. ID NO. 74 is the deduced amino acid sequence of TbRa28.
SEQ. ID NO. 75 is the deduced amino acid sequence of TbRa29.
SEQ. ID NO. 76 is the deduced amino acid sequence of TbRa2A.
5 SEQ. ID NO. 77 is the deduced amino acid sequence of TbRa3.
SEQ. ID NO. 78 is the deduced amino acid sequence of TbRa32.
SEQ. ID NO. 79 is the deduced amino acid sequence of TbRa35.
SEQ. ID NO. 80 is the deduced amino acid sequence of TbRa36.
SEQ. ID NO. 81 is the deduced amino acid sequence of TbRa4.
10 SEQ. ID NO. 82 is the deduced amino acid sequence of TbRa9.
SEQ. ID NO. 83 is the deduced amino acid sequence of TbRaB.
SEQ. ID NO. 84 is the deduced amino acid sequence of TbRaC.
SEQ. ID NO. 85 is the deduced amino acid sequence of TbRaD.
SEQ. ID NO. 86 is the deduced amino acid sequence of YYWCPG.
15 SEQ. ID NO. 87 is the deduced amino acid sequence of TbAAMK.
SEQ. ID NO. 88 is the deduced amino acid sequence of Tb38-1.
SEQ. ID NO. 89 is the deduced amino acid sequence of TbH-4.
SEQ. ID NO. 90 is the deduced amino acid sequence of TbH-8.
SEQ. ID NO. 91 is the deduced amino acid sequence of TbH-9.
20 SEQ. ID NO. 92 is the deduced amino acid sequence of TbH-12.
SEQ. ID NO. 93 is the amino acid sequence of Tb38-1 Peptide 1.
SEQ. ID NO. 94 is the amino acid sequence of Tb38-1 Peptide 2.
SEQ. ID NO. 95 is the amino acid sequence of Tb38-1 Peptide 3.
SEQ. ID NO. 96 is the amino acid sequence of Tb38-1 Peptide 4.
25 SEQ. ID NO. 97 is the amino acid sequence of Tb38-1 Peptide 5.
SEQ. ID NO. 98 is the amino acid sequence of Tb38-1 Peptide 6.
SEQ. ID NO. 99 is the DNA sequence of DPAS.
SEQ. ID NO. 100 is the deduced amino acid sequence of DPAS.

SEQ. ID NO. 103 is the DNA sequence of ESAT-6.

SEQ. ID NO. 104 is the deduced amino acid sequence of ESAT-6.

SEQ. ID NO. 105 is the DNA sequence of TbH-8-2.

SEQ. ID NO. 106 is the DNA sequence of TbH-9FL.

5 SEQ. ID NO. 107 is the deduced amino acid sequence of TbH-9FL.

SEQ. ID NO. 108 is the DNA sequence of TbH-9-1.

SEQ. ID NO. 109 is the deduced amino acid sequence of TbH-9-1.

SEQ. ID NO. 110 is the DNA sequence of TbH-9-4.

SEQ. ID NO. 111 is the deduced amino acid sequence of TbH-9-4.

10 SEQ. ID NO. 112 is the DNA sequence of Tb38-1F2 IN.

SEQ. ID NO. 113 is the DNA sequence of Tb38-2F2 RP.

SEQ. ID NO. 114 is the deduced amino acid sequence of Tb37-FL.

SEQ. ID NO. 115 is the deduced amino acid sequence of Tb38-IN.

SEQ. ID NO. 116 is the DNA sequence of Tb38-1F3.

15 SEQ. ID NO. 117 is the deduced amino acid sequence of Tb38-1F3.

SEQ. ID NO. 118 is the DNA sequence of Tb38-1F5.

SEQ. ID NO. 119 is the DNA sequence of Tb38-1F6.

SEQ. ID NO. 120 is the deduced N-terminal amino acid sequence of DPV.

SEQ. ID NO. 121 is the deduced N-terminal amino acid sequence of AVGS.

20 SEQ. ID NO. 122 is the deduced N-terminal amino acid sequence of AAMK.

SEQ. ID NO. 123 is the deduced N-terminal amino acid sequence of YYWC.

SEQ. ID NO. 124 is the deduced N-terminal amino acid sequence of DIGS.

SEQ. ID NO. 125 is the deduced N-terminal amino acid sequence of AEES.

SEQ. ID NO. 126 is the deduced N-terminal amino acid sequence of DPEP.

25 SEQ. ID NO. 127 is the deduced N-terminal amino acid sequence of APKT.

SEQ. ID NO. 128 is the deduced amino acid sequence of DPAS.

SEQ. ID NO. 129 is the protein sequence of DPPD N-terminal Antigen.

SEQ. ID NO. 130-133 are the protein sequences of four DPPD cyanogen

SEQ ID NO. 135 is the N-terminal protein sequence of AGD antigen.

SEQ ID NO. 136 is the N-terminal protein sequence of APE antigen.

SEQ ID NO. 137 is the N-terminal protein sequence of XYI antigen.

SEQ ID NO. 138 is the DNA sequence of TbH-29.

5 SEQ ID NO. 139 is the DNA sequence of TbH-30.

SEQ ID NO. 140 is the DNA sequence of TbH-32.

SEQ ID NO. 141 is the DNA sequence of TbH-33.

SEQ ID NO. 142 is the predicted amino acid sequence of TbH-29.

SEQ ID NO. 143 is the predicted amino acid sequence of TbH-30.

10 SEQ ID NO. 144 is the predicted amino acid sequence of TbH-32.

SEQ ID NO. 145 is the predicted amino acid sequence of TbH-33.

SEQ ID NO: 146-151 are PCR primers used in the preparation of a fusion protein containing TbRa3, 38 kD and Tb38-1.

15 SEQ ID NO: 152 is the DNA sequence of the fusion protein containing TbRa3, 38 kD and Tb38-1.

SEQ ID NO: 153 is the amino acid sequence of the fusion protein containing TbRa3, 38 kD and Tb38-1.

SEQ ID NO: 154 is the DNA sequence of the *M. tuberculosis* antigen 38 kD.

20 SEQ ID NO: 155 is the amino acid sequence of the *M. tuberculosis* antigen 38 kD.

SEQ ID NO: 156 is the DNA sequence of XP14.

SEQ ID NO: 157 is the DNA sequence of XP24

SEQ ID NO: 158 is the DNA sequence of XP31

SEQ ID NO: 159 is the 5' DNA sequence of XP32.

25 SEQ ID NO: 160 is the 3' DNA sequence of XP32.

SEQ ID NO: 161 is the predicted amino acid sequence of XP14.

SEQ ID NO: 162 is the predicted amino acid sequence encoded by the reverse complement of XP14.

SEQ ID NO: 163 is the

- SEQ ID NO: 165 is the 5' DNA sequence of XP4.
SEQ ID NO: 166 is the 5' DNA sequence of XP5.
SEQ ID NO: 167 is the 5' DNA sequence of XP17.
SEQ ID NO: 168 is the 5' DNA sequence of XP30.
5 SEQ ID NO: 169 is the 5' DNA sequence of XP2.
SEQ ID NO: 170 is the 3' DNA sequence of XP2.
SEQ ID NO: 171 is the 5' DNA sequence of XP3.
SEQ ID NO: 172 is the 3' DNA sequence of XP3.
SEQ ID NO: 173 is the 5' DNA sequence of XP6.
10 SEQ ID NO: 174 is the 3' DNA sequence of XP6.
SEQ ID NO: 175 is the 5' DNA sequence of XP18.
SEQ ID NO: 176 is the 3' DNA sequence of XP18.
SEQ ID NO: 177 is the 5' DNA sequence of XP19.
SEQ ID NO: 178 is the 3' DNA sequence of XP19.
15 SEQ ID NO: 179 is the 5' DNA sequence of XP22.
SEQ ID NO: 180 is the 3' DNA sequence of XP22.
SEQ ID NO: 181 is the 5' DNA sequence of XP25.
SEQ ID NO: 182 is the 3' DNA sequence of XP25.
SEQ ID NO: 183 is the full-length DNA sequence of TbH4-XP1.
20 SEQ ID NO: 184 is the predicted amino acid sequence of TbH4-XP1.
SEQ ID NO: 185 is the predicted amino acid sequence encoded by the reverse
complement of TbH4-XP1.
SEQ ID NO: 186 is a first predicted amino acid sequence encoded by XP36.
SEQ ID NO: 187 is a second predicted amino acid sequence encoded by XP36.
25 SEQ ID NO: 188 is the predicted amino acid sequence encoded by the reverse
complement of XP36.
SEQ ID NO: 189 is the DNA sequence of RDIF2.
SEQ ID NO: 190 is the DNA sequence of RDIF5.

SEQ ID NO: 193 is the DNA sequence of RDIF11.
SEQ ID NO: 194 is the predicted amino acid sequence of RDIF2.
SEQ ID NO: 195 is the predicted amino acid sequence of RDIF5.
SEQ ID NO: 196 is the predicted amino acid sequence of RDIF8.
5 SEQ ID NO: 197 is the predicted amino acid sequence of RDIF10.
SEQ ID NO: 198 is the predicted amino acid sequence of RDIF11.
SEQ ID NO: 199 is the 5' DNA sequence of RDIF12.
SEQ ID NO: 200 is the 3' DNA sequence of RDIF12.
SEQ ID NO: 201 is the DNA sequence of RDIF7.
10 SEQ ID NO: 202 is the predicted amino acid sequence of RDIF7.
SEQ ID NO: 203 is the DNA sequence of DIF2-1.
SEQ ID NO: 204 is the predicted amino acid sequence of DIF2-1.
SEQ ID NO: 205-212 are PCR primers used in the preparation of a fusion
protein containing TbRa3, 38 kD, Tb38-1 and DPEP (hereinafter referred to as
15 TbF-2).
SEQ ID NO: 213 is the DNA sequence of the fusion protein TbF-2.
SEQ ID NO: 214 is the amino acid sequence of the fusion protein TbF-2.
SEQ ID NO: 215 is the 5' DNA sequence of MO-1.
SEQ ID NO: 216 is the 5' DNA sequence for MO-2.
20 SEQ ID NO: 217 is the 5' DNA sequence for MO-4.
SEQ ID NO: 218 is the 5' DNA sequence for MO-8.
SEQ ID NO: 219 is the 5' DNA sequence for MO-9.
SEQ ID NO: 220 is the 5' DNA sequence for MO-26.
SEQ ID NO: 221 is the 5' DNA sequence for MO-28.
25 SEQ ID NO: 222 is the 5' DNA sequence for MO-29.
SEQ ID NO: 223 is the 5' DNA sequence for MO-30.
SEQ ID NO: 224 is the 5' DNA sequence for MO-34.
SEQ ID NO: 225 is the 5' DNA sequence for MO-35.

- SEQ ID NO: 228 is the predicted amino acid sequence for MO-4.
SEQ ID NO: 229 is the predicted amino acid sequence for MO-8.
SEQ ID NO: 230 is the predicted amino acid sequence for MO-9.
SEQ ID NO: 231 is the predicted amino acid sequence for MO-26.
5 SEQ ID NO: 232 is the predicted amino acid sequence for MO-28.
SEQ ID NO: 233 is the predicted amino acid sequence for MO-29.
SEQ ID NO: 234 is the predicted amino acid sequence for MO-30.
SEQ ID NO: 235 is the predicted amino acid sequence for MO-34.
SEQ ID NO: 236 is the predicted amino acid sequence for MO-35.
10 SEQ ID NO: 237 is the determined DNA sequence for MO-10.
SEQ ID NO: 238 is the predicted amino acid sequence for MO-10.
SEQ ID NO: 239 is the 3' DNA sequence for MO-27.
SEQ ID NO: 240 is the full-length DNA sequence for DPPD.
SEQ ID NO: 241 is the predicted full-length amino acid sequence for DPPD.
15 SEQ ID NO: 242 is the determined 5' cDNA sequence for LSER-10.
SEQ ID NO: 243 is the determined 5' cDNA sequence for LSER-11.
SEQ ID NO: 244 is the determined 5' cDNA sequence for LSER-12.
SEQ ID NO: 245 is the determined 5' cDNA sequence for LSER-13.
SEQ ID NO: 246 is the determined 5' cDNA sequence for LSER-16.
20 SEQ ID NO: 247 is the determined 5' cDNA sequence for LSER-25.
SEQ ID NO: 248 is the predicted amino acid sequence for LSER-10.
SEQ ID NO: 249 is the predicted amino acid sequence for LSER-12.
SEQ ID NO: 250 is the predicted amino acid sequence for LSER-13.
SEQ ID NO: 251 is the predicted amino acid sequence for LSER-16.
25 SEQ ID NO: 252 is the predicted amino acid sequence for LSER-25.
SEQ ID NO: 253 is the determined cDNA sequence for LSER-18.
SEQ ID NO: 254 is the determined cDNA sequence for LSER-23.
SEQ ID NO: 255 is the determined cDNA sequence for LSER-24.

SEQ ID NO: 258 is the predicted amino acid sequence for LSER-23
SEQ ID NO: 259 is the predicted amino acid sequence for LSER-24
SEQ ID NO: 260 is the predicted amino acid sequence for LSER-27
SEQ ID NO: 261 is the determined 5' cDNA sequence for LSER-1
5 SEQ ID NO: 262 is the determined 5' cDNA sequence for LSER-3
SEQ ID NO: 263 is the determined 5' cDNA sequence for LSER-4
SEQ ID NO: 264 is the determined 5' cDNA sequence for LSER-5
SEQ ID NO: 265 is the determined 5' cDNA sequence for LSER-6
SEQ ID NO: 266 is the determined 5' cDNA sequence for LSER-8
10 SEQ ID NO: 267 is the determined 5' cDNA sequence for LSER-14
SEQ ID NO: 268 is the determined 5' cDNA sequence for LSER-15
SEQ ID NO: 269 is the determined 5' cDNA sequence for LSER-17
SEQ ID NO: 270 is the determined 5' cDNA sequence for LSER-19
SEQ ID NO: 271 is the determined 5' cDNA sequence for LSER-20
15 SEQ ID NO: 272 is the determined 5' cDNA sequence for LSER-22
SEQ ID NO: 273 is the determined 5' cDNA sequence for LSER-26
SEQ ID NO: 274 is the determined 5' cDNA sequence for LSER-28
SEQ ID NO: 275 is the determined 5' cDNA sequence for LSER-29
SEQ ID NO: 276 is the determined 5' cDNA sequence for LSER-30
20 SEQ ID NO: 277 is the predicted amino acid sequence for LSER-1
SEQ ID NO: 278 is the predicted amino acid sequence for LSER-3
SEQ ID NO: 279 is the predicted amino acid sequence for LSER-5
SEQ ID NO: 280 is the predicted amino acid sequence for LSER-6
SEQ ID NO: 281 is the predicted amino acid sequence for LSER-8
25 SEQ ID NO: 282 is the predicted amino acid sequence for LSER-14
SEQ ID NO: 283 is the predicted amino acid sequence for LSER-15
SEQ ID NO: 284 is the predicted amino acid sequence for LSER-17
SEQ ID NO: 285 is the predicted amino acid sequence for LSER-19

SEQ ID NO: 288 is the predicted amino acid sequence for LSER-26

SEQ ID NO: 289 is the predicted amino acid sequence for LSER-28

SEQ ID NO: 290 is the predicted amino acid sequence for LSER-29

SEQ ID NO: 291 is the predicted amino acid sequence for LSER-30

5 SEQ ID NO: 292 is the determined cDNA sequence for LSER-9

SEQ ID NO: 293 is the determined cDNA sequence for the reverse complement of LSER-6

SEQ ID NO: 294 is the predicted amino acid sequence for the reverse complement of LSER-6

10 SEQ ID NO: 295 is the determined 5' cDNA sequence for MO-12

SEQ ID NO: 296 is the determined 5' cDNA sequence for MO-13

SEQ ID NO: 297 is the determined 5' cDNA sequence for MO-19

SEQ ID NO: 298 is the determined 5' cDNA sequence for MO-39

SEQ ID NO: 299 is the predicted amino acid sequence for MO-12

15 SEQ ID NO: 300 is the predicted amino acid sequence for MO-13

SEQ ID NO: 301 is the predicted amino acid sequence for MO-19

SEQ ID NO: 302 is the predicted amino acid sequence for MO-39

SEQ ID NO: 303 is the determined 5' cDNA sequence for Erdsn-1

SEQ ID NO: 304 is the determined 5' cDNA sequence for Erdsn-2

20 SEQ ID NO: 305 is the determined 5' cDNA sequence for Erdsn-4

SEQ ID NO: 306 is the determined 5' cDNA sequence for Erdsn-5

SEQ ID NO: 307 is the determined 5' cDNA sequence for Erdsn-6

SEQ ID NO: 308 is the determined 5' cDNA sequence for Erdsn-7

SEQ ID NO: 309 is the determined 5' cDNA sequence for Erdsn-8

25 SEQ ID NO: 310 is the determined 5' cDNA sequence for Erdsn-9

SEQ ID NO: 311 is the determined 5' cDNA sequence for Erdsn-10

SEQ ID NO: 312 is the determined 5' cDNA sequence for Erdsn-12

SEQ ID NO: 313 is the determined 5' cDNA sequence for Erdsn-13

SEQ ID NO: 314 is the determined 5' cDNA sequence for Erdsn-14

SEQ ID NO: 316 is the determined 5' cDNA sequence for Erdsn-16

SEQ ID NO: 317 is the determined 5' cDNA sequence for Erdsn-17

SEQ ID NO: 318 is the determined 5' cDNA sequence for Erdsn-18

SEQ ID NO: 319 is the determined 5' cDNA sequence for Erdsn-21

5 SEQ ID NO: 320 is the determined 5' cDNA sequence for Erdsn-22

SEQ ID NO: 321 is the determined 5' cDNA sequence for Erdsn-23

SEQ ID NO: 322 is the determined 5' cDNA sequence for Erdsn-25

SEQ ID NO: 323 is the determined 3' cDNA sequence for Erdsn-1

SEQ ID NO: 324 is the determined 3' cDNA sequence for Erdsn-2

10 SEQ ID NO: 325 is the determined 3' cDNA sequence for Erdsn-4

SEQ ID NO: 326 is the determined 3' cDNA sequence for Erdsn-5

SEQ ID NO: 327 is the determined 3' cDNA sequence for Erdsn-7

SEQ ID NO: 328 is the determined 3' cDNA sequence for Erdsn-8

SEQ ID NO: 329 is the determined 3' cDNA sequence for Erdsn-9

15 SEQ ID NO: 330 is the determined 3' cDNA sequence for Erdsn-10

SEQ ID NO: 331 is the determined 3' cDNA sequence for Erdsn-12

SEQ ID NO: 332 is the determined 3' cDNA sequence for Erdsn-13

SEQ ID NO: 333 is the determined 3' cDNA sequence for Erdsn-14

SEQ ID NO: 334 is the determined 3' cDNA sequence for Erdsn-15

20 SEQ ID NO: 335 is the determined 3' cDNA sequence for Erdsn-16

SEQ ID NO: 336 is the determined 3' cDNA sequence for Erdsn-17

SEQ ID NO: 337 is the determined 3' cDNA sequence for Erdsn-18

SEQ ID NO: 338 is the determined 3' cDNA sequence for Erdsn-21

SEQ ID NO: 339 is the determined 3' cDNA sequence for Erdsn-22

25 SEQ ID NO: 340 is the determined 3' cDNA sequence for Erdsn-23

SEQ ID NO: 341 is the determined 3' cDNA sequence for Erdsn-25

SEQ ID NO: 342 is the determined cDNA sequence for Erdsn-24

SEQ ID NO: 343 is the determined amino acid sequence for a *M. tuberculosis*

85b premature hemolysin

- SEQ ID NO: 345 is a determined amino acid sequence for spot 2
SEQ ID NO: 346 is a determined amino acid sequence for spot 2
SEQ ID NO: 347 is the determined amino acid seq for spot 4
SEQ ID NO: 348 is the sequence of primer PDM-157
5 SEQ ID NO: 349 is the sequence of primer PDM-160
SEQ ID NO: 350 is the DNA sequence of the fusion protein TbF-6
SEQ ID NO: 351 is the amino acid sequence of fusion protein TbF-6
SEQ ID NO: 352 is the sequence of primer PDM-176
SEQ ID NO: 353 is the sequence of primer PDM-175
10 SEQ ID NO: 354 is the DNA sequence of the fusion protein TbF-8
SEQ ID NO: 355 is the amino acid sequence of the fusion protein TbF-8

DETAILED DESCRIPTION OF THE INVENTION

As noted above, the present invention is generally directed to
15 compositions and methods for preventing, treating and diagnosing tuberculosis. The
compositions of the subject invention include polypeptides that comprise at least one
immunogenic portion of a *M. tuberculosis* antigen, or a variant of such an antigen that
differs only in conservative substitutions and/or modifications. Polypeptides within the
scope of the present invention include, but are not limited to, immunogenic soluble
20 *M. tuberculosis* antigens. A "soluble *M. tuberculosis* antigen" is a protein of
M. tuberculosis origin that is present in *M. tuberculosis* culture filtrate. As used herein,
the term "polypeptide" encompasses amino acid chains of any length, including full
length proteins (i.e., antigens), wherein the amino acid residues are linked by covalent
peptide bonds. Thus, a polypeptide comprising an immunogenic portion of one of the
25 above antigens may consist entirely of the immunogenic portion, or may contain
additional sequences. The additional sequences may be derived from the native
M. tuberculosis antigen or may be heterologous, and such sequences may (but need not)
be immunogenic.

of such antigens) are capable of stimulating cell proliferation, interleukin-12 production and/or interferon- γ production in biological samples comprising one or more cells selected from the group of T cells, NK cells, B cells and macrophages, where the cells are derived from an *M. tuberculosis*-immune individual. Polypeptides comprising at least an immunogenic portion of one or more *M. tuberculosis* antigens may generally be used to detect tuberculosis or to induce protective immunity against tuberculosis in a patient.

The compositions and methods of the present invention also encompass variants of the above polypeptides and DNA molecules. A polypeptide "variant," as used herein, is a polypeptide that differs from the recited polypeptide only in conservative substitutions and/or modifications, such that the therapeutic, antigenic and/or immunogenic properties of the polypeptide are retained. Polypeptide variants preferably exhibit at least about 70%, more preferably at least about 90% and most preferably at least about 95% identity to the identified polypeptides. For polypeptides with immunoreactive properties, variants may, alternatively, be identified by modifying the amino acid sequence of one of the above polypeptides, and evaluating the immunoreactivity of the modified polypeptide. For polypeptides useful for the generation of diagnostic binding agents, a variant may be identified by evaluating a modified polypeptide for the ability to generate antibodies that detect the presence or absence of tuberculosis. Such modified sequences may be prepared and tested using, for example, the representative procedures described herein.

As used herein, a "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydropathic nature of the polypeptide to be substantially unchanged. In general, the following groups of amino acids represent conservative changes: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his.

antigenic properties, secondary structure and hydrophobic nature of the polypeptide. For example, a polypeptide may be conjugated to a signal (or leader) sequence at the N-terminal end of the protein which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (e.g., poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

A nucleotide "variant" is a sequence that differs from the recited nucleotide sequence in having one or more nucleotide deletions, substitutions or additions. Such modifications may be readily introduced using standard mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis as taught, for example, by Adelman et al. (*DNA*, 2:183, 1983). Nucleotide variants may be naturally occurring allelic variants, or non-naturally occurring variants. Variant nucleotide sequences preferably exhibit at least about 70%, more preferably at least about 80% and most preferably at least about 90% identity to the recited sequence. Such variant nucleotide sequences will generally hybridize to the recited nucleotide sequence under stringent conditions. As used herein, "stringent conditions" refers to prewashing in a solution of 6X SSC, 0.2% SDS; hybridizing at 65 °C, 6X SSC, 0.2% SDS overnight; followed by two washes of 30 minutes each in 1X SSC, 0.1% SDS at 65 °C and two washes of 30 minutes each in 0.2X SSC, 0.1% SDS at 65 °C.

In a related aspect, combination polypeptides are disclosed. A "combination polypeptide" is a polypeptide comprising at least one of the above immunogenic portions and one or more additional immunogenic *M. tuberculosis* sequences, which are joined via a peptide linkage into a single amino acid chain. The sequences may be joined directly (i.e., with no intervening amino acids) or may be joined by way of a linker sequence (e.g., Gly-Cys-Gly) that does not significantly diminish the immunogenic properties of the component polypeptides.

In general, *M. tuberculosis* antigens, and DNA sequences encoding them,

those of ordinary skill in the art, including anion-exchange and reverse phase chromatography. Purified antigens are then evaluated for their ability to elicit an appropriate immune response (e.g., cellular) using, for example, the representative methods described herein. Immunogenic antigens may then be partially sequenced
5 using techniques such as traditional Edman chemistry. See Edman and Berg, *Eur. J. Biochem.* 80:116-132, 1967.

Immunogenic antigens may also be produced recombinantly using a DNA sequence that encodes the antigen, which has been inserted into an expression vector and expressed in an appropriate host. DNA molecules encoding soluble antigens
10 may be isolated by screening an appropriate *M. tuberculosis* expression library with anti-sera (e.g., rabbit) raised specifically against soluble *M. tuberculosis* antigens. DNA sequences encoding antigens that may or may not be soluble may be identified by screening an appropriate *M. tuberculosis* genomic or cDNA expression library with sera obtained from patients infected with *M. tuberculosis*. Such screens may generally be
15 performed using techniques well known to those of ordinary skill in the art, such as those described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989.

DNA sequences encoding soluble antigens may also be obtained by screening an appropriate *M. tuberculosis* cDNA or genomic DNA library for DNA
20 sequences that hybridize to degenerate oligonucleotides derived from partial amino acid sequences of isolated soluble antigens. Degenerate oligonucleotide sequences for use in such a screen may be designed and synthesized, and the screen may be performed, as described (for example) in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989 (and references cited therein). Polymerase chain reaction (PCR) may also be employed, using the above
25 oligonucleotides in methods well known in the art, to isolate a nucleic acid probe from a cDNA or genomic library. The library screen may then be performed using the isolated probe.

Alternatively, genomic or cDNA libraries may be screened using

lines or clones derived from one or more *M. tuberculosis*-immune individuals. In general, PBMCs and/or T cells for use in such screens may be prepared as described below. Direct library screens may generally be performed by assaying pools of expressed recombinant proteins for the ability to induce proliferation and/or interferon- γ production in T cells derived from an *M. tuberculosis*-immune individual. Alternatively, potential T cell antigens may be first selected based on antibody reactivity, as described above.

Regardless of the method of preparation, the antigens (and immunogenic portions thereof) described herein (which may or may not be soluble) have the ability to induce an immunogenic response. More specifically, the antigens have the ability to induce proliferation and/or cytokine production (i.e., interferon- γ and/or interleukin-12 production) in T cells, NK cells, B cells and/or macrophages derived from an *M. tuberculosis*-immune individual. The selection of cell type for use in evaluating an immunogenic response to a antigen will, of course, depend on the desired response. For example, interleukin-12 production is most readily evaluated using preparations containing B cells and/or macrophages. An *M. tuberculosis*-immune individual is one who is considered to be resistant to the development of tuberculosis by virtue of having mounted an effective T cell response to *M. tuberculosis* (i.e., substantially free of disease symptoms). Such individuals may be identified based on a strongly positive (i.e., greater than about 10 mm diameter induration) intradermal skin test response to tuberculosis proteins (PPD) and an absence of any signs or symptoms of tuberculosis disease. T cells, NK cells, B cells and macrophages derived from *M. tuberculosis*-immune individuals may be prepared using methods known to those of ordinary skill in the art. For example, a preparation of PBMCs (i.e., peripheral blood mononuclear cells) may be employed without further separation of component cells. PBMCs may generally be prepared, for example, using density centrifugation through FicollTM (Winthrop Laboratories, NY). T cells for use in the assays described herein may also be purified directly from PBMCs. Alternatively, an enriched T cell line reactive against mycobacterial proteins is available from the American Tissue Culture Collection.

PBMCs from *M. tuberculosis*-immune individuals with mycobacterial proteins for a period of 2-4 weeks. This allows expansion of only the mycobacterial protein-specific T cells, resulting in a line composed solely of such cells. These cells may then be cloned and tested with individual proteins, using methods known to those of ordinary skill in the art, to more accurately define individual T cell specificity. In general, antigens that test positive in assays for proliferation and/or cytokine production (*i.e.*, interferon- γ and/or interleukin-12 production) performed using T cells, NK cells, B cells and/or macrophages derived from an *M. tuberculosis*-immune individual are considered immunogenic. Such assays may be performed, for example, using the representative procedures described below. Immunogenic portions of such antigens may be identified using similar assays, and may be present within the polypeptides described herein.

The ability of a polypeptide (*e.g.*, an immunogenic antigen, or a portion or other variant thereof) to induce cell proliferation is evaluated by contacting the cells (*e.g.*, T cells and/or NK cells) with the polypeptide and measuring the proliferation of the cells. In general, the amount of polypeptide that is sufficient for evaluation of about 10^5 cells ranges from about 10 ng/mL to about 100 μ g/mL and preferably is about 10 μ g/mL. The incubation of polypeptide with cells is typically performed at 37°C for about six days. Following incubation with polypeptide, the cells are assayed for a proliferative response, which may be evaluated by methods known to those of ordinary skill in the art, such as exposing cells to a pulse of radiolabeled thymidine and measuring the incorporation of label into cellular DNA. In general, a polypeptide that results in at least a three fold increase in proliferation above background (*i.e.*, the proliferation observed for cells cultured without polypeptide) is considered to be able to induce proliferation.

The ability of a polypeptide to stimulate the production of interferon- γ and/or interleukin-12 in cells may be evaluated by contacting the cells with the polypeptide and measuring the level of interferon- γ or interleukin-12 produced by the cells. In general, the amount of polypeptide that is sufficient for the evaluation of about 10^5 cells ranges from about 10 ng/mL to about 100 μ g/mL and preferably is about 10 μ g/mL.

bead or a biodegradable microsphere, such as those described in U.S. Patent Nos. 4,897,268 and 5,075,109. The incubation of polypeptide with the cells is typically performed at 37°C for about six days. Following incubation with polypeptide, the cells are assayed for interferon- γ and/or interleukin-12 (or one or more subunits thereof), which may be evaluated by methods known to those of ordinary skill in the art, such as an enzyme-linked immunosorbent assay (ELISA) or, in the case of IL-12 P70 subunit, a bioassay such as an assay measuring proliferation of T cells. In general, a polypeptide that results in the production of at least 50 pg of interferon- γ per mL of cultured supernatant (containing 10^4 - 10^5 T cells per mL) is considered able to stimulate the production of interferon- γ . A polypeptide that stimulates the production of at least 10 pg/mL of IL-12 P70 subunit, and/or at least 100 pg/mL of IL-12 P40 subunit, per 10^5 macrophages or B cells (or per 3×10^5 PBMC) is considered able to stimulate the production of IL-12.

In general, immunogenic antigens are those antigens that stimulate proliferation and/or cytokine production (i.e., interferon- γ and/or interleukin-12 production) in T cells, NK cells, B cells and/or macrophages derived from at least about 25% of *M. tuberculosis*-immune individuals. Among these immunogenic antigens, polypeptides having superior therapeutic properties may be distinguished based on the magnitude of the responses in the above assays and based on the percentage of individuals for which a response is observed. In addition, antigens having superior therapeutic properties will not stimulate proliferation and/or cytokine production *in vitro* in cells derived from more than about 25% of individuals that are not *M. tuberculosis*-immune, thereby eliminating responses that are not specifically due to *M. tuberculosis*-responsive cells. Those antigens that induce a response in a high percentage of T cell, NK cell, B cell and/or macrophage preparations from *M. tuberculosis*-immune individuals (with a low incidence of responses in cell preparations from other individuals) have superior therapeutic properties.

Antigens with superior therapeutic properties may also be identified based on their ability to stimulate a response in

for use on experimental animals are described in detail below. Efficacy may be determined based on the ability of the antigen to provide at least about a 50% reduction in bacterial numbers and/or at least about a 40% decrease in mortality following experimental infection. Suitable experimental animals include mice, guinea pigs and
5 primates.

Antigens having superior diagnostic properties may generally be identified based on the ability to elicit a response in an intradermal skin test performed on an individual with active tuberculosis, but not in a test performed on an individual who is not infected with *M. tuberculosis*. Skin tests may generally be performed as
10 described below, with a response of at least 5 mm induration considered positive.

Immunogenic portions of the antigens described herein may be prepared and identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3d ed., Raven Press, 1993, pp. 243-247 and references cited therein. Such techniques include screening polypeptide portions of the native antigen
15 for immunogenic properties. The representative proliferation and cytokine production assays described herein may generally be employed in these screens. An immunogenic portion of a polypeptide is a portion that, within such representative assays, generates an immune response (e.g., proliferation, interferon- γ production and/or interleukin-12 production) that is substantially similar to that generated by the full length antigen. In
20 other words, an immunogenic portion of an antigen may generate at least about 20%, and preferably about 100%, of the proliferation induced by the full length antigen in the model proliferation assay described herein. An immunogenic portion may also, or alternatively, stimulate the production of at least about 20%, and preferably about 100%, of the interferon- and/or interleukin-12 induced by the full length antigen in the
25 model assay described herein.

Portions and other variants of *M. tuberculosis* antigens may be generated by synthetic or recombinant means. Synthetic polypeptides having fewer than about 100 amino acids, and generally fewer than about 50 amino acids, may be generated using techniques well known in the art.

techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. See Merrifield, *J. Am. Chem. Soc.* 85:2149-2146, 1963. Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Applied BioSystems, Inc., Foster City, CA, and may be operated according to the manufacturer's instructions. Variants of a native antigen may generally be prepared using standard mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis. Sections of the DNA sequence may also be removed using standard techniques to permit preparation of truncated polypeptides.

Recombinant polypeptides containing portions and/or variants of a native antigen may be readily prepared from a DNA sequence encoding the polypeptide using a variety of techniques well known to those of ordinary skill in the art. For example, supernatants from suitable host/vector systems which secrete recombinant protein into culture media may be first concentrated using a commercially available filter. Following concentration, the concentrate may be applied to a suitable purification matrix such as an affinity matrix or an ion exchange resin. Finally, one or more reverse phase HPLC steps can be employed to further purify a recombinant protein.

Any of a variety of expression vectors known to those of ordinary skill in the art may be employed to express recombinant polypeptides of this invention. Expression may be achieved in any appropriate host cell that has been transformed or transfected with an expression vector containing a DNA molecule that encodes a recombinant polypeptide. Suitable host cells include prokaryotes, yeast and higher eukaryotic cells. Preferably, the host cells employed are *E. coli*, yeast or a mammalian cell line such as COS or CHO. The DNA sequences expressed in this manner may encode naturally occurring antigens, portions of naturally occurring antigens, or other variants thereof.

In general, regardless of the method of preparation, the polypeptides disclosed herein are substantially pure.

preferably at least about 99% pure. In certain preferred embodiments, described in detail below, the substantially pure polypeptides are incorporated into pharmaceutical compositions or vaccines for use in one or more of the methods disclosed herein.

In certain specific embodiments, the subject invention discloses polypeptides comprising at least an immunogenic portion of a soluble *M. tuberculosis* antigen having one of the following N-terminal sequences, or a variant thereof that differs only in conservative substitutions and/or modifications:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu; (SEQ ID No. 120)
- (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser; (SEQ ID No. 121)
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg; (SEQ ID No. 122)
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro; (SEQ ID No. 123)
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val; (SEQ ID No. 124)
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro; (SEQ ID No. 125)
- (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Ala-Ala-Ala-Ser-Pro-Pro-Ser; (SEQ ID No. 126)
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly; (SEQ ID No. 127)
- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn; (SEQ ID No. 128)
- (j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID No. 129)
- (k) Xaa-Ile-Val-Thr-Ser-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID No. 130)

(f) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 136)

wherein Xaa may be any amino acid, preferably a cysteine residue. A DNA sequence encoding the antigen identified as (g) above is provided in SEQ ID No. 52, and the polypeptide encoded by SEQ ID No. 52 is provided in SEQ ID No. 53. A DNA sequence encoding the antigen defined as (a) above is provided in SEQ ID No. 101; its deduced amino acid sequence is provided in SEQ ID No. 102. A DNA sequence corresponding to antigen (d) above is provided in SEQ ID No. 24 a DNA sequence corresponding to antigen (c) is provided in SEQ ID No. 25 and a DNA sequence corresponding to antigen (i) is provided in SEQ ID No. 99; its deduced amino acid sequence is provided in SEQ ID No. 100.

In a further specific embodiment, the subject invention discloses polypeptides comprising at least an immunogenic portion of an *M. tuberculosis* antigen having one of the following N-terminal sequences, or a variant thereof that differs only in conservative substitutions and/or modifications:

(m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID No. 137) or

(n) Asp-Pro-Pro-Asp-Pro-Ile-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Xaa-Phe; (SEQ ID No. 129)

wherein Xaa may be any amino acid, preferably a cysteine residue.

In other specific embodiments, the subject invention discloses polypeptides comprising at least an immunogenic portion of a soluble *M. tuberculosis* antigen (or a variant of such an antigen) that comprises one or more of the amino acid sequences encoded by (a) the DNA sequences of SEQ ID Nos. 1, 2, 4-10, 13-25 and 52; (b) the complements of such DNA sequences, or (c) DNA sequences substantially homologous to a sequence in (a) or (b).

In further specific embodiments, the subject invention discloses polypeptides comprising at least an immunogenic portion of a *M. tuberculosis* antigen (or a variant of such an antigen) which may or may not be soluble, that comprises

Nos.: 26-51, 138, 139, 163-183, 189-193, 199, 200, 201, 203, 215-225, 239, 240, 242-247, 253-256, 261-276, 292, 293, 295-298 and 303-342. (b) the complements of such DNA sequences or (c) DNA sequences substantially homologous to a sequence in (a) or (b).

5 In the specific embodiments discussed above, the *M. tuberculosis* antigens include variants that are encoded by DNA sequences which are substantially homologous to one or more of DNA sequences specifically recited herein. "Substantial homology," as used herein, refers to DNA sequences that are capable of hybridizing under moderately stringent conditions. Suitable moderately stringent conditions include
10 prewashing in a solution of 5X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50°C-65°C, 5X SSC, overnight or, in the case of cross-species homology at 45°C, 0.5X SSC; followed by washing twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS). Such hybridizing DNA sequences are also within the scope of this invention, as are nucleotide sequences that, due to code
15 degeneracy, encode an immunogenic polypeptide that is encoded by a hybridizing DNA sequence.

In a related aspect, the present invention provides fusion proteins comprising a first and a second inventive polypeptide or, alternatively, a polypeptide of the present invention and a known *M. tuberculosis* antigen, such as the 38 kD antigen
20 described in Andersen and Hansen, *Infect. Immun.* 57:2481-2488, 1989, (Genbank Accession No. M30046) or ESAT-6 (SEQ ID Nos. 103 and 104), together with variants of such fusion proteins. The fusion proteins of the present invention may also include a linker peptide between the first and second polypeptides.

A DNA sequence encoding a fusion protein of the present invention is
25 constructed using known recombinant DNA techniques to assemble separate DNA sequences encoding the first and second polypeptides into an appropriate expression vector. The 3' end of a DNA sequence encoding the first polypeptide is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide so that the reading frames of the sequences are in phase to permit mRNA

translation of the two DNA sequences into a single fusion protein that retains the biological activity of both the first and the second polypeptides.

A peptide linker sequence may be employed to separate the first and the second polypeptides by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion protein using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea et al., *Gene* 40:39-46, 1985; Murphy et al., *Proc. Natl. Acad. Sci. USA* 83:8258-8262, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may be from 1 to about 50 amino acids in length. Peptide sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

The ligated DNA sequences are operably linked to suitable transcriptional or translational regulatory elements. The regulatory elements responsible for expression of DNA are located only 5' to the DNA sequence encoding the first polypeptides. Similarly, stop codons require to end translation and transcription termination signals are only present 3' to the DNA sequence encoding the second polypeptide.

In another aspect, the present invention provides methods for using one or more of the above polypeptides or fusion proteins (or DNA molecules encoding such polypeptides) to induce protective immunity against tuberculosis in a patient. As used herein, a "patient" refers to any mammal, including humans.

may be afflicted with a disease, or may be free of detectable disease and/or infection. In other words, protective immunity may be induced to prevent or treat tuberculosis.

In this aspect, the polypeptide, fusion protein or DNA molecule is generally present within a pharmaceutical composition and/or a vaccine. Pharmaceutical compositions may comprise one or more polypeptides, each of which may contain one or more of the above sequences (or variants thereof), and a physiologically acceptable carrier. Vaccines may comprise one or more of the above polypeptides and a non-specific immune response enhancer, such as an adjuvant or a liposome (into which the polypeptide is incorporated). Such pharmaceutical compositions and vaccines may also contain other *M. tuberculosis* antigens, either incorporated into a combination polypeptide or present within a separate polypeptide.

Alternatively, a vaccine may contain DNA encoding one or more polypeptides as described above, such that the polypeptide is generated *in situ*. In such vaccines, the DNA may be present within any of a variety of delivery systems known to those of ordinary skill in the art, including nucleic acid expression systems, bacterial and viral expression systems. Appropriate nucleic acid expression systems contain the necessary DNA sequences for expression in the patient (such as a suitable promoter and terminating signal). Bacterial delivery systems involve the administration of a bacterium (such as *Bacillus Calmette-Guérin*) that expresses an immunogenic portion of the polypeptide on its cell surface. In a preferred embodiment, the DNA may be introduced using a viral expression system (e.g., vaccinia or other pox virus, retrovirus, or adenovirus), which may involve the use of a non-pathogenic (defective), replication competent virus. Techniques for incorporating DNA into such expression systems are well known to those of ordinary skill in the art. The DNA may also be "naked," as described, for example, in Ulmer et al., *Science* 259:1745-1749, 1993 and reviewed by Cohen, *Science* 259:1691-1692, 1993. The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells.

In a related aspect, the present invention provides a method for

invention or a known *M. tuberculosis* antigen, such as the 38 kD antigen described above. For example, administration of DNA encoding a polypeptide of the present invention, either "naked" or in a delivery system as described above, may be followed by administration of an antigen in order to enhance the protective immune effect of the vaccine.

Routes and frequency of administration, as well as dosage, will vary from individual to individual and may parallel those currently being used in immunization using BCG. In general, the pharmaceutical compositions and vaccines may be administered by injection (e.g., intracutaneous, intramuscular, intravenous or subcutaneous), intranasally (e.g., by aspiration) or orally. Between 1 and 3 doses may be administered for a 1-36 week period. Preferably, 3 doses are administered, at intervals of 3-4 months, and booster vaccinations may be given periodically thereafter. Alternate protocols may be appropriate for individual patients. A suitable dose is an amount of polypeptide or DNA that, when administered as described above, is capable of raising an immune response in an immunized patient sufficient to protect the patient from *M. tuberculosis* infection for at least 1-2 years. In general, the amount of polypeptide present in a dose (or produced *in situ* by the DNA in a dose) ranges from about 1 pg to about 100 mg per kg of host, typically from about 10 pg to about 1 mg, and preferably from about 100 pg to about 1 μ g. Suitable dose sizes will vary with the size of the patient, but will typically range from about 0.1 mL to about 5 mL.

While any suitable carrier known to those of ordinary skill in the art may be employed in the pharmaceutical compositions of this invention, the type of carrier will vary depending on the mode of administration. For parenteral administration, such as subcutaneous injection, the carrier preferably comprises water, saline, alcohol, a fat, a wax or a buffer. For oral administration, any of the above carriers or a solid carrier, such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, talcum, cellulose, glucose, sucrose, and magnesium carbonate, may be employed. Biodegradable microspheres (e.g., polylactic acid/lactide) may also be employed as carriers for the pharmaceutical compositions of this invention.

Any of a variety of adjuvants may be employed in the vaccines of this invention to nonspecifically enhance the immune response. Most adjuvants contain a substance designed to protect the antigen from rapid catabolism, such as aluminum hydroxide or mineral oil, and a nonspecific stimulator of immune responses, such as lipid A, *Bordetella pertussis* or *Mycobacterium tuberculosis*. Suitable adjuvants are commercially available as, for example, Freund's Incomplete Adjuvant and Freund's Complete Adjuvant (Difco Laboratories) and Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ). Other suitable adjuvants include alum, biodegradable microspheres, monophosphoryl lipid A and quil A.

In another aspect, this invention provides methods for using one or more of the polypeptides described above to diagnose tuberculosis using a skin test. As used herein, a "skin test" is any assay performed directly on a patient in which a delayed-type hypersensitivity (DTH) reaction (such as swelling, reddening or dermatitis) is measured following intradermal injection of one or more polypeptides as described above. Such injection may be achieved using any suitable device sufficient to contact the polypeptide or polypeptides with dermal cells of the patient, such as a tuberculin syringe or 1 mL syringe. Preferably, the reaction is measured at least 48 hours after injection, more preferably 48-72 hours.

The DTH reaction is a cell-mediated immune response, which is greater in patients that have been exposed previously to the test antigen (*i.e.*, the immunogenic portion of the polypeptide employed, or a variant thereof). The response may be measured visually, using a ruler. In general, a response that is greater than about 0.5 cm in diameter, preferably greater than about 1.0 cm in diameter, is a positive response, indicative of tuberculosis infection, which may or may not be manifested as an active disease.

The polypeptides of this invention are preferably formulated, for use in a skin test, as pharmaceutical compositions containing a polypeptide and a physiologically acceptable carrier, as described above. Such compositions typically contain one or more of the above polypeptides in an amount ranging from about 0.1 mg to 10 mg.

Preferably, the carrier employed in such pharmaceutical compositions is a saline solution with appropriate preservatives, such as phenol and/or Tween 80™.

In a preferred embodiment, a polypeptide employed in a skin test is of sufficient size such that it remains at the site of injection for the duration of the reaction period. In general, a polypeptide that is at least 9 amino acids in length is sufficient. The polypeptide is also preferably broken down by macrophages within hours of injection to allow presentation to T-cells. Such polypeptides may contain repeats of one or more of the above sequences and/or other immunogenic or nonimmunogenic sequences.

The following Examples are offered by way of illustration and not by way of limitation.

EXAMPLES

EXAMPLE 1

PURIFICATION AND CHARACTERIZATION OF POLYPEPTIDES FROM *M. TUBERCULOSIS* CULTURE FILTRATE

This example illustrates the preparation of *M. tuberculosis* soluble polypeptides from culture filtrate. Unless otherwise noted, all percentages in the following example are weight per volume.

M. tuberculosis (either H37Ra, ATCC No. 25177, or H37Rv, ATCC No. 25618) was cultured in sterile GAS media at 37°C for fourteen days. The media was then vacuum filtered (leaving the bulk of the cells) through a 0.45 μ filter into a sterile 2.5 L bottle. The media was next filtered through a 0.2 μ filter into a sterile 4 L bottle and NaN₃ was added to the culture filtrate to a concentration of 0.04%. The bottles were then placed in a 4°C cold room.

The culture filtrate was concentrated by ultrafiltration.

which had been rinsed with ethanol and contained a 10,000 kDa MWCO membrane. The pressure was maintained at 60 psi using nitrogen gas. This procedure reduced the 12 L volume to approximately 50 ml.

The culture filtrate was dialyzed into 0.1% ammonium bicarbonate using
5 a 8,000 kDa MWCO cellulose ester membrane, with two changes of ammonium bicarbonate solution. Protein concentration was then determined by a commercially available BCA assay (Pierce, Rockford, IL).

The dialyzed culture filtrate was then lyophilized, and the polypeptides resuspended in distilled water. The polypeptides were dialyzed against 0.01 mM 1,3
10 bis[tris(hydroxymethyl)-methylamino]propane, pH 7.5 (Bis-Tris propane buffer), the initial conditions for anion exchange chromatography. Fractionation was performed using gel perfusion chromatography on a POROS 146 II Q/M anion exchange column 4.6 mm x 100 mm (Perseptive BioSystems, Framingham, MA) equilibrated in 0.01 mM Bis-Tris propane buffer pH 7.5. Polypeptides were eluted with a linear 0-0.5 M NaCl
15 gradient in the above buffer system. The column eluent was monitored at a wavelength of 220 nm.

The pools of polypeptides eluting from the ion exchange column were dialyzed against distilled water and lyophilized. The resulting material was dissolved in
0.1% trifluoroacetic acid (TFA) pH 1.9 in water, and the polypeptides were purified on
20 a Delta-Pak C18 column (Waters, Milford, MA) 300 Angstrom pore size, 5 micron particle size (3.9 x 150 mm). The polypeptides were eluted from the column with a linear gradient from 0-60% dilution buffer (0.1% TFA in acetonitrile). The flow rate was 0.75 ml/minute and the HPLC eluent was monitored at 214 nm. Fractions containing the eluted polypeptides were collected to maximize the purity of the
25 individual samples. Approximately 200 purified polypeptides were obtained.

The purified polypeptides were then screened for the ability to induce T-cell proliferation in PBMC preparations. The PBMCs from donors known to be PPD skin test positive and whose T-cells were shown to proliferate in response to PPD and crude soluble proteins from MTH were used for this purpose.

polypeptides were added in duplicate at concentrations of 0.5 to 10 $\mu\text{g/mL}$. After six days of culture in 96-well round-bottom plates in a volume of 200 μl , 50 μl of medium was removed from each well for determination of IFN- γ levels, as described below. The plates were then pulsed with 1 $\mu\text{Ci/well}$ of tritiated thymidine for a further 18 hours, harvested and tritium uptake determined using a gas scintillation counter. Fractions that resulted in proliferation in both replicates three fold greater than the proliferation observed in cells cultured in medium alone were considered positive.

IFN- γ was measured using an enzyme-linked immunosorbent assay (ELISA). ELISA plates were coated with a mouse monoclonal antibody directed to human IFN- γ (PharMingen, San Diego, CA) in PBS for four hours at room temperature. Wells were then blocked with PBS containing 5% (W/V) non-fat dried milk for 1 hour at room temperature. The plates were then washed six times in PBS/0.2% TWEEN-20 and samples diluted 1:2 in culture medium in the ELISA plates were incubated overnight at room temperature. The plates were again washed and a polyclonal rabbit anti-human IFN- γ serum diluted 1:3000 in PBS/10% normal goat serum was added to each well. The plates were then incubated for two hours at room temperature, washed and horseradish peroxidase-coupled anti-rabbit IgG (Sigma Chemical Co., St. Louis, MO) was added at a 1:2000 dilution in PBS/5% non-fat dried milk. After a further two hour incubation at room temperature, the plates were washed and TMB substrate added. The reaction was stopped after 20 min with 1 N sulfuric acid. Optical density was determined at 450 nm using 570 nm as a reference wavelength. Fractions that resulted in both replicates giving an OD two fold greater than the mean OD from cells cultured in medium alone, plus 3 standard deviations, were considered positive.

For sequencing, the polypeptides were individually dried onto Biobrene MB (Perkin Elmer/Applied BioSystems Division, Foster City, CA) treated glass fiber filters. The filters with polypeptide were loaded onto a Perkin Elmer/Applied BioSystems Division Procise 492 protein sequencer. The polypeptides were sequenced from the amino terminal and using traditional Edman chemistry. The amino acid sequence was determined for each polypeptide.

Using the procedure described above, antigens having the following N-terminal sequences were isolated:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Xaa-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu; (SEQ ID No. 54)
- (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser; (SEQ ID No. 55)
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg; (SEQ ID No. 56)
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro; (SEQ ID No. 57)
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val; (SEQ ID No. 58)
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro; (SEQ ID No. 59)
- (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Ala-Ala-Ala-Ala-Pro-Pro-Ala; (SEQ ID No. 60) and
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly; (SEQ ID No. 61)

wherein Xaa may be any amino acid.

An additional antigen was isolated employing a microbore HPLC purification step in addition to the procedure described above. Specifically, 20 μ l of a fraction comprising a mixture of antigens from the chromatographic purification step previously described, was purified on an Aquapore C18 column (Perkin Elmer Applied Biosystems Division, Foster City, CA) with a 7 micron pore size, column size 1 mm x 100 mm, in a Perkin Elmer Applied Biosystems Division Model 172 HPLC. Fractions were eluted from the column with a linear gradient of 1% minute of acetonitrile (containing 0.05% TFA) in water (0.05% TFA) at a flow rate of 80 μ l/minute. The eluent was monitored at 250 nm. The original fraction was separated into 4 major peaks plus other smaller component and was analyzed by mass spectrometry.

have a molecular weight of 12.054 Kd (by mass spectrometry) and the following N-terminal sequence:

- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Gln-
 Thr-Ser-Leu-Leu-Asn-Asn-Leu-Ala-Asp-Pro-Asp-Val-Ser-Phe-
 5 Ala-Asp (SEQ ID No. 62).

This polypeptide was shown to induce proliferation and IFN- γ production in PBMC preparations using the assays described above.

Additional soluble antigens were isolated from *M. tuberculosis* culture filtrate as follows. *M. tuberculosis* culture filtrate was prepared as described above. Following dialysis against Bis-Tris propane buffer, at pH 5.5, fractionation was
 10 performed using anion exchange chromatography on a Poros QE column 4.6 x 100 mm (Perseptive Biosystems) equilibrated in Bis-Tris propane buffer pH 5.5. Polypeptides were eluted with a linear 0-1.5 M NaCl gradient in the above buffer system at a flow rate of 10 ml/min. The column eluent was monitored at a wavelength of 214 nm.

The fractions eluting from the ion exchange column were pooled and
 15 subjected to reverse phase chromatography using a Poros R2 column 4.6 x 100 mm (Perseptive Biosystems). Polypeptides were eluted from the column with a linear gradient from 0-100% acetonitrile (0.1% TFA) at a flow rate of 5 ml/min. The eluent was monitored at 214 nm.

Fractions containing the eluted polypeptides were lyophilized and
 20 resuspended in 80 μ l of aqueous 0.1% TFA and further subjected to reverse phase chromatography on a Vydac C4 column 4.6 x 150 mm (Western Analytical, Temecula, CA) with a linear gradient of 0-100% acetonitrile (0.1% TFA) at a flow rate of 2 ml/min. Eluent was monitored at 214 nm.

The fraction with biological activity was separated into one major peak
 25 plus other smaller components. Western blot of this peak onto PVDF membrane revealed three major bands of molecular weights 14 Kd, 20 Kd and 26 Kd. These polypeptides were determined to have the following N-terminal sequence, respectively

- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Gln-
 Thr-Ser-Leu-Leu-Asn-Asn-Leu-Ala-Asp-Pro-Asp-Val-Ser-Phe-
 Ala-Asp (SEQ ID No. 62).

(k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID No. 135) and

(l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 136), wherein Xaa may be any amino acid.

5 Using the assays described above, these polypeptides were shown to induce proliferation and IFN- γ production in PBMC preparations. Figs. 1A and B show the results of such assays using PBMC preparations from a first and a second donor, respectively.

DNA sequences that encode the antigens designated as (a), (c), (d) and
10 (g) above were obtained by screening a genomic *M. tuberculosis* library using ³²P end labeled degenerate oligonucleotides corresponding to the N-terminal sequence and containing *M. tuberculosis* codon bias. The screen performed using a probe corresponding to antigen (a) above identified a clone having the sequence provided in SEQ ID No. 101. The polypeptide encoded by SEQ ID No. 101 is provided in SEQ ID
15 No. 102. The screen performed using a probe corresponding to antigen (g) above identified a clone having the sequence provided in SEQ ID No. 52. The polypeptide encoded by SEQ ID No. 52 is provided in SEQ ID No. 53. The screen performed using a probe corresponding to antigen (d) above identified a clone having the sequence provided in SEQ ID No. 24, and the screen performed with a probe corresponding to
20 antigen (e) identified a clone having the sequence provided in SEQ ID No. 25.

The above amino acid sequences were compared to known amino acid sequences in the gene bank using the DNA STAR system. The database searched contains some 173,000 proteins and is a combination of the Swiss, PIR databases along with translated protein sequences (Version 87). No significant homologies to the amino
25 acid sequences for antigens (a)-(h) and (l) were detected.

The amino acid sequence for antigen (i) was found to be homologous to a sequence from *M. leprae*. The full length *M. leprae* sequence was amplified from genomic DNA using the sequence obtained from GENBANK. This sequence was then used to screen the *M. tuberculosis* library. The amino acid sequence for antigen (i) is provided in SEQ ID No. 137.

The amino acid sequence for antigen (j) was found to be homologous to a known *M. tuberculosis* protein translated from a DNA sequence. To the best of the inventors' knowledge, this protein has not been previously shown to possess T-cell stimulatory activity. The amino acid sequence for antigen (k) was found to be related to a sequence from *M. leprae*.

In the proliferation and IFN- γ assays described above, using three PPD positive donors, the results for representative antigens provided above are presented in Table 1:

TABLE 1
RESULTS OF PBMC PROLIFERATION AND IFN- γ ASSAYS

Sequence	Proliferation	IFN- γ
(a)	+	-
(c)	+++	+++
(d)	++	++
(g)	+++	+++
(h)	+++	+++

In Table 1, responses that gave a stimulation index (SI) of between 2 and 4 (compared to cells cultured in medium alone) were scored as +, an SI of 4-8 or 2-4 at a concentration of 1 μ g or less was scored as ++ and an SI of greater than 8 was scored as +++. The antigen of sequence (h) was found to have a high SI (++++) for one donor and lower SI (++ and +) for the two other donors in both proliferation and IFN- γ assays. These results indicate that these antigens are capable of inducing proliferation and/or interferon- γ production.

EXAMPLE 2USE OF PATIENT SERA TO ISOLATE *M. TUBERCULOSIS* ANTIGENS

This example illustrates the isolation of antigens from *M. tuberculosis* lysate by screening with serum from *M. tuberculosis*-infected individuals.

Dessicated *M. tuberculosis* H37Ra (Difco Laboratories) was added to a 2% NP40 solution, and alternately homogenized and sonicated three times. The resulting suspension was centrifuged at 13,000 rpm in microfuge tubes and the supernatant put through a 0.2 micron syringe filter. The filtrate was bound to Macro Prep DEAE beads (BioRad, Hercules, CA). The beads were extensively washed with 20 mM Tris pH 7.5 and bound proteins eluted with 1M NaCl. The 1M NaCl elute was dialyzed overnight against 10 mM Tris, pH 7.5. Dialyzed solution was treated with DNase and RNase at 0.05 mg/ml for 30 min. at room temperature and then with α -D-mannosidase, 0.5 U/mg at pH 4.5 for 3-4 hours at room temperature. After returning to pH 7.5, the material was fractionated via FPLC over a Bio Scale-Q-20 column (BioRad). Fractions were combined into nine pools, concentrated in a Centriprep 10 (Amicon, Beverley, MA) and then screened by Western blot for serological activity using a serum pool from *M. tuberculosis*-infected patients which was not immunoreactive with other antigens of the present invention.

The most reactive fraction was run in SDS-PAGE and transferred to PVDF. A band at approximately 85 Kd was cut out yielding the sequence:

(m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; SEQ ID No. 137), wherein Xaa may be any amino acid

Comparison of this sequence with those in the gene bank as described above, revealed no significant homologies to known sequences.

A DNA sequence that encodes the antigen designated as (m) above was obtained by screening a genomic *M. tuberculosis* Erdman strain library using labeled degenerate oligonucleotides corresponding to the amino acid sequence (m).

203. This sequence was found to encode the amino acid sequence provided in SEQ ID
NO: 204. Comparison of these sequences with those in the genebank revealed some
similarity to sequences previously identified in *M. tuberculosis* and *M. bovis*.

5

EXAMPLE 3

PREPARATION OF DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

This example illustrates the preparation of DNA sequences encoding
M. tuberculosis antigens by screening a *M. tuberculosis* expression library with sera
10 obtained from patients infected with *M. tuberculosis*, or with anti-sera raised against
soluble *M. tuberculosis* antigens.

A. PREPARATION OF *M. TUBERCULOSIS* SOLUBLE ANTIGENS USING RABBIT ANTI-SERA RAISED AGAINST *M. TUBERCULOSIS* SUPERNATANT

15 Genomic DNA was isolated from the *M. tuberculosis* strain H37Ra. The
DNA was randomly sheared and used to construct an expression library using the
Lambda ZAP expression system (Stratagene, La Jolla, CA). Rabbit anti-sera was
generated against secretory proteins of the *M. tuberculosis* strains H37Ra, H37Rv and
Erdman by immunizing a rabbit with concentrated supernatant of the *M. tuberculosis*
20 cultures. Specifically, the rabbit was first immunized subcutaneously with 200 µg of
protein antigen in a total volume of 2 ml containing 10 µg muramyl dipeptide
(Calbiochem, La Jolla, CA) and 1 ml of incomplete Freund's adjuvant. Four weeks later
the rabbit was boosted subcutaneously with 100 µg antigen in incomplete Freund's
adjuvant. Finally, the rabbit was immunized intravenously four weeks later with 50 µg
25 protein antigen. The anti-sera were used to screen the expression library as described in
Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor
Laboratories, Cold Spring Harbor, NY, 1989. Bacteriophage plaques expressing
immunoreactive antigens were purified. Phagemid from the plaques was rescued and
the nucleotide sequences of the *M. tuberculosis* clones deduced.

were expressed and purified antigens used in the immunological analysis described in Example 1. Proteins were induced by IPTG and purified by gel elution, as described in Skeiky et al., *J. Exp. Med.* 181:1527-1537, 1995. Representative sequences of DNA molecules identified in this screen are provided in SEQ ID Nos. 1-25. The
5 corresponding predicted amino acid sequences are shown in SEQ ID Nos. 63-87.

On comparison of these sequences with known sequences in the gene bank using the databases described above, it was found that the clones referred to hereinafter as TbRA2A, TbRA16, TbRA18, and TbRA29 (SEQ ID Nos. 76, 68, 70, 75) show some homology to sequences previously identified in *Mycobacterium leprae* but
10 not in *M. tuberculosis*. TbRA2A was found to be a lipoprotein, with a six residue lipidation sequence being located adjacent to a hydrophobic secretory sequence. TbRA11, TbRA26, TbRA28 and TbDPEP (SEQ ID Nos. 65, 73, 74, 53) have been previously identified in *M. tuberculosis*. No significant homologies were found to
15 TbRA1, TbRA3, TbRA4, TbRA9, TbRA10, TbRA13, TbRA17, TbRa19, TbRA29, TbRA32, TbRA36 and the overlapping clones TbRA35 and TbRA12 (SEQ ID Nos. 63, 77, 81, 82, 64, 67, 69, 71, 75, 78, 80, 79, 66). The clone TbRa24 is overlapping with clone TbRa29.

The results of PBMC proliferation and interferon- γ assays performed on representative recombinant antigens, and using T-cell preparations from several
20 different *M. tuberculosis*-immune patients, are presented in Tables 2 and 3, respectively.

TABLE 2
RESULTS OF PBMC PROLIFERATION TO REPRESENTATIVE SOLUBLE ANTIGENS

Antigen	Patient												
	1	2	3	4	5	6	7	8	9	10	11	12	13
IbRa1	-	+	+	+	-	-	+	+	-	-	+	+	-
IbRa3	-	+	+	-	+	-	-	++	+	-	-	+	-
IbRa9	-	-	nt	nt	++	++	nt	nt	nt	nt	nt	nt	nt
IbRa10	-	+	+	+	+	+	nt	+	-	+	+	nt	nt
IbRa11	+	+	+	+	++	+	nt	-	++	++	+	+	-
IbRa12	-	-	+	+	+	++	+	+	+	-	+	+	nt
IbRa16	nt	nt	nt	nt	-	+	nt	nt	nt	nt	nt	nt	-
IbRa24	nt	nt	nt	nt	-	-	nt	nt	nt	nt	nt	nt	nt
IbRa26	-	nt	nt	nt	-	-	nt	nt	nt	nt	nt	nt	nt
IbRa29	nt	nt	nt	nt	-	-	nt	nt	nt	nt	nt	nt	nt
IbRa35	-	nt	++	++	++	++	nt	++	++	++	++	++	nt
IbRaB	nt	nt	nt	nt	-	-	nt	nt	nt	nt	nt	nt	nt
IbRaC	nt	nt	nt	nt	-	-	nt	nt	nt	nt	nt	nt	nt
IbRaD	nt	nt	nt	nt	-	-	nt	nt	nt	nt	nt	nt	nt
AAMK	-	-	+	-	-	-	nt	nt	nt	nt	nt	nt	nt
YY	-	-	-	-	-	-	nt	-	-	-	nt	+	nt
DTP-P	-	-	-	-	-	-	nt	++	+	+	nt	+	nt
Control	-	-	-	-	-	-	-	-	-	-	-	-	-

nt tested

TABLE 3
RESULTS OF PBMC CELL IRON- γ PRODUCTION TO REPRESENTATIVE SOLUBLE ANTIGENS

Antigen	Patient												
	1	2	3	4	5	6	7	8	9	10	11	12	13
HLRA1	+	+	+	+	+	-	-	+	-	-	+	+	-
HLRA3	+	+	+	+	+	-	-	+	-	-	-	+	-
HLRA9	+	+	+	+	+	-	-	+	-	-	-	-	-
HLRA10	+	+	+	+	+	-	-	+	-	-	-	-	-
HLRA11	+	+	+	+	+	-	-	+	-	-	-	-	-
HLRA12	+	+	+	+	+	-	-	+	-	-	-	-	-
HLRA16	+	+	+	+	+	-	-	+	-	-	-	-	-
HLRA24	+	+	+	+	+	-	-	+	-	-	-	-	-
HLRA26	+	+	+	+	+	-	-	+	-	-	-	-	-
HLRA29	+	+	+	+	+	-	-	+	-	-	-	-	-
HLRA35	+	+	+	+	+	-	-	+	-	-	-	-	-
HLRA3	+	+	+	+	+	-	-	+	-	-	-	-	-
HLRA4	+	+	+	+	+	-	-	+	-	-	-	-	-
HLRA4D	+	+	+	+	+	-	-	+	-	-	-	-	-
AAAIK	+	+	+	+	+	-	-	+	-	-	-	-	-
YY	+	+	+	+	+	-	-	+	-	-	-	-	-
DPEP	+	+	+	+	+	-	-	+	-	-	-	-	-
Control	+	+	+	+	+	-	-	+	-	-	-	-	-

In Tables 2 and 3, responses that gave a stimulation index (SI) of between 1.2 and 2 (compared to cells cultured in medium alone) were scored as \pm , a SI of 2-4 was scored as +, as SI of 4-8 or 2-4 at a concentration of 1 μ g or less was scored as ++ and an SI of greater than 8 was scored as ++++. In addition, the effect of concentration on proliferation and interferon- γ production is shown for two of the above antigens in the attached Figure. For both proliferation and interferon- γ production, TbRa3 was scored as ++ and TbRa9 as +.

These results indicate that these soluble antigens can induce proliferation and/or interferon- γ production in T-cells derived from an *M. tuberculosis*-immune individual.

B. USE OF SERA FROM PATIENTS HAVING PULMONARY OR PLEURAL TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

The genomic DNA library described above, and an additional H37Rv library, were screened using pools of sera obtained from patients with active tuberculosis. To prepare the H37Rv library, *M. tuberculosis* strain H37Rv genomic DNA was isolated, subjected to partial *Sau3A* digestion and used to construct an expression library using the Lambda Zap expression system (Stratagene, La Jolla, Cal). Three different pools of sera, each containing sera obtained from three individuals with active pulmonary or pleural disease, were used in the expression screening. The pools were designated TbL, TbM and TbH, referring to relative reactivity with H37Ra lysate (i.e., TbL = low reactivity, TbM = medium reactivity and TbH = high reactivity) in both ELISA and immunoblot format. A fourth pool of sera from seven patients with active pulmonary tuberculosis was also employed. All of the sera lacked increased reactivity with the recombinant 38 kD *M. tuberculosis* H37Ra phosphate-binding protein.

All pools were pre-adsorbed with *E. coli* lysate and used to screen the H37Ra and H37Rv expression libraries, as described in Sambrook et al., *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989. Bacteriophage plaques expressing immunoreactive antigens were purified

Thirty two clones were purified. Of these, 31 represented sequences that had not been previously identified in human *M. tuberculosis*. Representative sequences of the DNA molecules identified are provided in SEQ ID Nos.: 26-51 and 105. Of these, TbH-8-2 (SEQ. ID NO. 105) is a partial clone of TbH-8, and TbH-4 (SEQ. ID NO. 43) and TbH-4-FWD (SEQ. ID NO. 44) are non-contiguous sequences from the same clone. Amino acid sequences for the antigens hereinafter identified as Tb38-1, TbH-4, TbH-8, TbH-9, and TbH-12 are shown in SEQ ID Nos.: 88-92. Comparison of these sequences with known sequences in the gene bank using the databases identified above revealed no significant homologies to TbH-4, TbH-8, TbH-9 and TbM-3, although weak homologies were found to TbH-9. TbH-12 was found to be homologous to a 34 kD antigenic protein previously identified in *M. paratuberculosis* (Acc. No. S28515). Tb38-1 was found to be located 34 base pairs upstream of the open reading frame for the antigen ESAT-6 previously identified in *M. bovis* (Acc. No. U34848) and in *M. tuberculosis* (Sorensen et al., *Infect. Immun.* 63:1710-1717, 1995).


Probes derived from Tb38-1 and TbH-9, both isolated from an H37Ra library, were used to identify clones in an H37Rv library. Tb38-1 hybridized to Tb38-1F2, Tb38-1F3, Tb38-1F5 and Tb38-1F6 (SEQ. ID NOS. 112, 113, 116, 118, and 119). (SEQ ID NOS. 112 and 113 are non-contiguous sequences from clone Tb38-1F2). Two open reading frames were deduced in Tb38-1F2: one corresponds to Tb37FL (SEQ. ID. NO. 114), the second, a partial sequence, may be the homologue of Tb38-1 and is called Tb38-IN (SEQ. ID NO. 115). The deduced amino acid sequence of Tb38-1F3 is presented in SEQ. ID. NO. 117. A TbH-9 probe identified three clones in the H37Rv library: TbH-9-FL (SEQ. ID NO. 106), which may be the homologue of TbH-9 (H37Ra), TbH-9-1 (SEQ. ID NO. 108), and TbH-9-4 (SEQ. ID NO. 110), all of which are highly related sequences to TbH-9. The deduced amino acid sequences for these three clones are presented in SEQ ID NOS. 107, 109 and 111.

Further screening of the *M. tuberculosis* genomic DNA library, as described above, resulted in the identification of additional clones.

above, one was determined to be identical to the 14Kd alpha crystallin heat shock protein previously shown to be present in *M. tuberculosis*, and a third was determined to be identical to the antigen TbH-8 described above. The determined DNA sequences for the remaining five clones (hereinafter referred to as TbH-29, TbH-30, TbH-32 and 5 TbH-33) are provided in SEQ ID NO: 138-141, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 142-145, respectively. The DNA and amino acid sequences for these antigens were compared with those in the gene bank as described above. No homologies were found to the 5' end of TbH-29, (which contains the reactive open reading frame), although the 3' end of TbH-29 10 was found to be identical to the *M. tuberculosis* cosmid Y227. TbH-32 and TbH-33 were found to be identical to the previously identified *M. tuberculosis* insertion element IS6110 and to the *M. tuberculosis* cosmid Y50, respectively. No significant homologies to TbH-30 were found.

Positive phagemid from this additional screening were used to infect *E. coli* XL-1 Blue MRF', as described in Sambrook et al., *supra*. Induction of recombinant 15 protein was accomplished by the addition of IPTG. Induced and uninduced lysates were run in duplicate on SDS-PAGE and transferred to nitrocellulose filters. Filters were reacted with human *M. tuberculosis* sera (1:200 dilution) reactive with TbH and a rabbit sera (1:200 or 1:250 dilution) reactive with the N-terminal 4 Kd portion of lacZ. 20 Sera incubations were performed for 2 hours at room temperature. Bound antibody was detected by addition of ¹²⁵I-labeled Protein A and subsequent exposure to film for variable times ranging from 16 hours to 11 days. The results of the immunoblots are summarized in Table 4

TABLE 4

<u>Antigen</u>	<u>Human M. tb Sera</u>	<u>Anti-lacZ Sera</u>
TbH-29	45 Kd	45 Kd
TbH-30	No reactivity	29 Kd
TbH-32	12 Kd	12 Kd
TbH-33	16 Kd 	16 Kd

10

Positive reaction of the recombinant human *M. tuberculosis* antigens with both the human *M. tuberculosis* sera and anti-lacZ sera indicate that reactivity of the human *M. tuberculosis* sera is directed towards the fusion protein. Antigens reactive with the anti-lacZ sera but not with the human *M. tuberculosis* sera may be the result of the human *M. tuberculosis* sera recognizing conformational epitopes, or the antigen-antibody binding kinetics may be such that the 2 hour sera exposure in the immunoblot is not sufficient.

The results of T-cell assays performed on Tb38-1, ESAT-6 and other representative recombinant antigens are presented in Tables 5A, B and C, respectively, below:

TABLE 5A
RESULTS OF PBMC PROLIFERATION TO REPRESENTATIVE ANTIGENS

Antigen	Donor										
	1	2	3	4	5	6	7	8	9	10	11
Tb38-1	—	+	+	+	+	+	+	+	+	+	+
ESAT-6	—	+	+	+	+	+	+	+	+	+	+
TbH-29	—	—	—	—	—	—	—	—	—	—	—

15

TABLE 5B
RESULTS OF PBMC INTERFERON- γ PRODUCTION TO REPRESENTATIVE ANTIGENS

Antigen	Donor										
	1	2	3	4	5	6	7	8	9	10	11
Tb38.1	+++	+	-	+	+	+++	-	++	-	+++	+++
ESAT-6	+++	+	-	+	++	+	-	+	-	+++	+++
TbH-9	++	++	-	+++	±	±	+++	+++	++	+++	++

5

TABLE 6
SUMMARY OF T-CELL RESPONSES TO REPRESENTATIVE ANTIGENS

Antigen	Proliferation			Interferon- γ			total
	patient 4	patient 5	patient 6	patient 4	patient 5	patient 6	
TbH9	++	++	++	+++	++	++	13
TbM7	-	-	-	++	+	-	4
TbH5	-	+	+	++	++	++	8
TbL23	-	+	=	++	+++	+	7.5
TbH4	-	+++	±	+++	+++	±	7
- control	-	-	-	-	-	-	0

10

These results indicate that both the inventive *M. tuberculosis* antigens and ESAT-6 can induce proliferation and/or interferon- γ production in T-cells derived from an *M. tuberculosis*-immune individual. To the best of the inventors' knowledge, ESAT-6 has not been previously shown to stimulate human immune responses

A set of six overlapping peptides covering the amino acid sequence of the antigen Tb38-1 was constructed using the method described in Example 6. The sequences of these peptides, hereinafter referred to as pep1-6, are provided in SEQ ID Nos. 93-98, respectively. The results of T-cell assays using these peptides are shown in Tables 7 and 8. These results confirm the existence, and help to localize T-cell epitopes within Tb38-1 capable of inducing proliferation and interferon- γ production in T-cells.

15

TABLE 7
RESULTS OF PBMC PROLIFERATION TO TB38.1 PEPTIDES

[illegible]

Studies were undertaken to determine whether the antigens TbH-9 and Tb38-1 represent cellular proteins or are secreted into *M. tuberculosis* culture media. In the first study, rabbit sera were raised against A) secretory proteins of *M. tuberculosis*, B) the known secretory recombinant *M. tuberculosis* antigen 85b, C) recombinant Tb38-1 and D) recombinant TbH-9, using protocols substantially the same as that as described in Example 3A. Total *M. tuberculosis* lysate, concentrated supernatant of *M. tuberculosis* cultures and the recombinant antigens 85b, TbH-9 and Tb38-1 were resolved on denaturing gels, immobilized on nitrocellulose membranes and duplicate blots were probed using the rabbit sera described above.

The results of this analysis using control sera (panel I) and antisera (panel II) against secretory proteins, recombinant 85b, recombinant Tb38-1 and recombinant TbH-9 are shown in Figures 3A-D, respectively, wherein the lane designations are as follows: 1) molecular weight protein standards; 2) 5 μ g of *M. tuberculosis* lysate; 3) 5 μ g secretory proteins; 4) 50 ng recombinant Tb38-1; 5) 50 ng recombinant TbH-9; and 6) 50 ng recombinant 85b. The recombinant antigens were engineered with six terminal histidine residues and would therefore be expected to migrate with a mobility approximately 1 kD larger than the native protein. In Figure 3D, recombinant TbH-9 is lacking approximately 10 kD of the full-length 42 kD antigen, hence the significant difference in the size of the immunoreactive native TbH-9 antigen in the lysate lane (indicated by an arrow). These results demonstrate that Tb38-1 and TbH-9 are intracellular antigens and are not actively secreted by *M. tuberculosis*.

The finding that TbH-9 is an intracellular antigen was confirmed by determining the reactivity of TbH-9-specific human T cell clones to recombinant TbH-9, secretory *M. tuberculosis* proteins and PPD. A TbH-9-specific T cell clone (designated 131TbH-9) was generated from PBMC of a healthy PPD-positive donor. The proliferative response of 131TbH-9 to secretory proteins, recombinant TbH-9 and a control *M. tuberculosis* antigen, TbRa11, was determined by measuring uptake of tritiated thymidine, as described in Example 1. As shown in Figure 4A, the clone 131TbH-9 responds specifically to TbH-9, showing that TbH-9 is not a significant component of the secretory proteins.

(designated PPD 800-10) prepared from PBMC from a healthy PPD-positive donor, following stimulation of the T cell clone with secretory proteins, PPD or recombinant TbH-9. These results further confirm that TbH-9 is not secreted by *M. tuberculosis*.

C. USE OF SERA FROM PATIENTS HAVING EXTRAPULMONARY TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

Genomic DNA was isolated from *M. tuberculosis* Erdman strain, randomly sheared and used to construct an expression library employing the Lambda ZAP expression system (Stratagene, La Jolla, CA). The resulting library was screened using pools of sera obtained from individuals with extrapulmonary tuberculosis, as described above in Example 3B, with the secondary antibody being goat anti-human IgG + A + M (H+L) conjugated with alkaline phosphatase.

Eighteen clones were purified. Of these, 4 clones (hereinafter referred to as XP14, XP24, XP31 and XP32) were found to bear some similarity to known sequences. The determined DNA sequences for XP14, XP24 and XP31 are provided in SEQ ID Nos.: 156-158, respectively, with the 5' and 3' DNA sequences for XP32 being provided in SEQ ID Nos.: 159 and 160, respectively. The predicted amino acid sequence for XP14 is provided in SEQ ID No.: 161. The reverse complement of XP14 was found to encode the amino acid sequence provided in SEQ ID No.: 162.

Comparison of the sequences for the remaining 14 clones (hereinafter referred to as XP1-XP6, XP17-XP19, XP22, XP25, XP27, XP30 and XP36) with those in the genbank as described above, revealed no homologies with the exception of the 3' ends of XP2 and XP6 which were found to bear some homology to known *M. tuberculosis* cosmid. The DNA sequences for XP27 and XP36 are shown in SEQ ID Nos.: 163 and 164, respectively, with the 5' sequences for XP4, XP5, XP17 and XP30 being shown in SEQ ID Nos.: 165-168, respectively, and the 5' and 3' sequences for XP2, XP3, XP6, XP18, XP19, XP22 and XP25 being shown in SEQ ID Nos.: 169 and 170; 171 and 172; 173 and 174; 175 and 176; 177 and 178; 179 and 180; and 181 and 182, respectively. XP1 was found to

open reading frame encoding the amino acid sequence shown in SEQ ID No: 184. The reverse complement of TbH4-XP1 was found to contain an open reading frame encoding the amino acid sequence shown in SEQ ID No.: 185. The DNA sequence for XP36 was found to contain two open reading frames encoding the amino acid sequence shown in SEQ ID Nos.: 186 and 187, with the reverse complement containing an open reading frame encoding the amino acid sequence shown in SEQ ID No.: 188.

Recombinant XP1 protein was prepared as described above in Example 3B, with a metal ion affinity chromatography column being employed for purification. As illustrated in Figures 8A-B and 9A-B, using the assays described herein, recombinant XP1 was found to stimulate cell proliferation and IFN- γ production in T cells isolated from an *M. tuberculosis*-immune donors.

D. USE OF A LYSATE POSITIVE SERUM POOL FROM PATIENTS HAVING TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

Genomic DNA was isolated from *M. tuberculosis* Erdman strain, randomly sheared and used to construct an expression library employing the Lambda Screen expression system (Novagen, Madison, WI), as described below in Example 6. Pooled serum obtained from *M. tuberculosis*-infected patients and that was shown to react with *M. tuberculosis* lysate but not with the previously expressed proteins 38kD, Tb38-1, TbRa3, TbH4, DPEP and TbRa11, was used to screen the expression library as described above in Example 3B, with the secondary antibody being goat anti-human IgG - A - M (H+L) conjugated with alkaline phosphatase.

Twenty-seven clones were purified. Comparison of the determined cDNA sequences for these clones revealed no significant homologies to 10 of the clones (hereinafter referred to as LSER-10, LSER-11, LSER-12, LSER-13, LSER-16, LSER-18, LSER-23, LSER-24, LSER-25 and LSER-27). The determined cDNA sequences for LSER-10, LSER-11, LSER-12, LSER-13, LSER-16 and LSER-25 are provided in SEQ ID Nos. 242-247, respectively, with the corresponding predicted amino acid sequences for LSER-10,

and LSER-27 are shown in SEQ ID NO: 253-256, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 257-260. The remaining seventeen clones were found to show similarities to unknown sequences previously identified in *M. tuberculosis*. The determined 5' cDNA sequences for sixteen of these clones (hereinafter referred to as LSER-1, LSER-3, LSER-4, LSER-5, LSER-6, LSER-8, LSER-14, LSER-15, LSER-17, LSER-19, LSER-20, LSER-22, LSER-26, LSER-28, LSER-29 and LSER-30) are provided in SEQ ID NO: 261-276, respectively, with the corresponding predicted amino acid sequences for LSER-1, LSER-3, LSER-5, LSER-6, LSER-8, LSER-14, LSER-15, LSER-17, LSER-19, LSER-20, LSER-22, LSER-26, LSER-28, LSER-29 and LSER-30 being provided in SEQ ID NO: 277-291, respectively. The determined full-length cDNA sequence for the clone LSER-9 is provided in SEQ ID NO: 292. The reverse complement of LSER-6 (SEQ ID NO: 293) was found to encode the predicted amino acid sequence of SEQ ID NO: 294.

E. PREPARATION OF *M. TUBERCULOSIS* SOLUBLE ANTIGENS USING RABBIT ANTI-SERA RAISED AGAINST *M. TUBERCULOSIS* FRACTIONATED PROTEINS

M. tuberculosis lysate was prepared as described above in Example 2. The resulting material was fractionated by HPLC and the fractions screened by Western blot for serological activity with a serum pool from *M. tuberculosis*-infected patients which showed little or no immunoreactivity with other antigens of the present invention. Rabbit anti-sera was generated against the most reactive fraction using the method described in Example 3A. The anti-sera was used to screen an *M. tuberculosis* Erdman strain genomic DNA expression library prepared as described above. Bacteriophage plaques expressing immunoreactive antigens were purified. Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones determined.

Ten different clones were purified. Of these, one was found to be TbRa35, described above, and one was found to be the previously identified *M. tuberculosis* antigen, HSP60. Of the remaining eight clones, seven (hereinafter referred to as RDIF2, RDIF5, RDIF8, RDIF10, RDIF11 and RDIF12) were found to bear some similarity to previously

corresponding predicted amino acid sequences being provided in SEQ ID Nos: 194-198, respectively. The 5' and 3' DNA sequences for RDIF12 are provided in SEQ ID Nos.: 199 and 200, respectively. No significant homologies were found to the antigen RDIF-7. The determined DNA and predicted amino acid sequences for RDIF7 are provided in SEQ ID Nos.: 201 and 202, respectively. One additional clone, referred to as RDIF6 was isolated, however, this was found to be identical to RDIF5.

Recombinant RDIF6, RDIF8, RDIF10 and RDIF11 were prepared as described above. As shown in Figures 8A-B and 9A-B, these antigens were found to stimulate cell proliferation and IFN- γ production in T cells isolated from *M. tuberculosis*-immune donors.

EXAMPLE 4

PURIFICATION AND CHARACTERIZATION OF A POLYPEPTIDE FROM TUBERCULIN PURIFIED PROTEIN DERIVATIVE

An *M. tuberculosis* polypeptide was isolated from tuberculin purified protein derivative (PPD) as follows.

PPD was prepared as published with some modification (Seibert, F. et al., Tuberculin purified protein derivative. Preparation and analyses of a large quantity for standard. The American Review of Tuberculosis 44:9-25, 1941).

M. tuberculosis Rv strain was grown for 6 weeks in synthetic medium in roller bottles at 37°C. Bottles containing the bacterial growth were then heated to 100° C in water vapor for 3 hours. Cultures were sterile filtered using a 0.22 μ filter and the liquid phase was concentrated 20 times using a 3 kD cut-off membrane. Proteins were precipitated once with 50% ammonium sulfate solution and eight times with 25% ammonium sulfate solution. The resulting proteins (PPD) were fractionated by reverse phase liquid chromatography (RP HPLC) using a C18 column (7.8 x 300 mM; Waters, Milford, MA) in a Biocad HPLC system (Perseptive Biosystems, Framingham, MA). Fractions were eluted from the column with a linear gradient from 0-100% buffer (0.1% TFA in acetonitrile). The flow rate was 1.0 ml/minute and eluent was monitored at 214 nm and 280 nm.

Six fractions were collected, dried, suspended in PBS and tested individually in *M. tuberculosis*-infected guinea pigs for induction of delayed type hypersensitivity (DTH) reaction. One fraction was found to induce a strong DTH reaction and was subsequently fractionated further by RP-HPLC on a microbore Vydac C18 column (Cat. No. 218TP5115) in a Perkin Elmer/Applied Biosystems Division Model 172 HPLC. Fractions were eluted with a linear gradient from 5-100% buffer (0.05% TFA in acetonitrile) with a flow rate of 80 μ l/minute. Eluent was monitored at 215 nm. Eight fractions were collected and tested for induction of DTH in *M. tuberculosis*-infected guinea pigs. One fraction was found to induce strong DTH of about 16 mm induration. The other fractions did not induce detectable DTH. The positive fraction was submitted to SDS-PAGE gel electrophoresis and found to contain a single protein band of approximately 12 kD molecular weight.

This polypeptide, herein after referred to as DPPD, was sequenced from the amino terminal using a Perkin Elmer/Applied Biosystems Division Procise 492 protein sequencer as described above and found to have the N-terminal sequence shown in SEQ ID No.: 129. Comparison of this sequence with known sequences in the gene bank as described above revealed no known homologies. Four cyanogen bromide fragments of DPPD were isolated and found to have the sequences shown in SEQ ID Nos.: 130-133. A subsequent search of the *M. tuberculosis* genome database released by the Institute for Genomic Research revealed a match of the DPPD partial amino acid sequence with a sequence present within the *M. tuberculosis* cosmid MTY21C12. An open reading frame of 336 bp was identified. The full-length DNA sequence for DPPD is provided in SEQ ID NO: 240, with the corresponding full-length amino acid sequence being provided in SEQ ID NO: 241.

The ability of the antigen DPPD to stimulate human PBMC to proliferate and to produce IFN- γ was assayed as described in Example 1. As shown in Table 9, DPPD was found to stimulate proliferation and elicit production of large quantities of IFN- γ , more than that elicited by commercial PPD.

TABLE 9
RESULTS OF PROLIFERATION AND INTERFERON- γ ASSAYS TO DPPD

PBMC Donor	Stimulator	Proliferation (CPM)	IFN- γ (OD ₄₅₀)
A	Medium	1,089	0.17
	PPD (commercial)	8,394	1.29
	DPPD	13,451	2.21
B	Medium	450	0.09
	PPD (commercial)	3,929	1.26
	DPPD	6,184	1.49
C	Medium	541	0.11
	PPD (commercial)	8,907	0.76
	DPPD	23,024	>2.70

EXAMPLE 5

USE OF SERA FROM TUBERCULOSIS-INFECTED MONKEYS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

Genomic DNA was isolated from *M. tuberculosis* Erdman strain, randomly sheared and used to construct an expression library employing the Lambda ZAP expression system (Stratagene, La Jolla, CA). Serum samples were obtained from a cynomolgous monkey 18, 33, 51 and 56 days following infection with *M. tuberculosis* Erdman strain. These samples were pooled and used to screen the *M. tuberculosis* genomic DNA expression library using the procedure described above in Example 3C.

Twenty clones were purified. The determined 5' DNA sequences for the clones referred to as MO-1, MO-2, MO-4, MO-8, MO-9, MO-26, MO-28, MO-29, MO-30, MO-34 and MO-35 are provided in SEQ ID NO: 213-225, respectively, with the corresponding

predicted amino acid sequence being provided in SEQ ID NO: 238. The 3' DNA sequence for the clone MO-27 is provided in SEQ ID NO: 239.

Clones MO-1, MO-30 and MO-35 were found to show a high degree of relatedness and showed some homology to a previously identified unknown *M. tuberculosis* sequence and to cosmid MTC1237. MO-2 was found to show some homology to aspartokinase from *M. tuberculosis*. Clones MO-3, MO-7 and MO-27 were found to be identical and to show a high degree of relatedness to MO-5. All four of these clones showed some homology to *M. tuberculosis* heat shock protein 70. MO-27 was found to show some homology to *M. tuberculosis* cosmid MTCY339. MO-4 and MO-34 were found to show some homology to cosmid SCY21B4 and *M. smegmatis* integration host factor, and were both found to show some homology to a previously identified, unknown *M. tuberculosis* sequence. MO-6 was found to show some homology to *M. tuberculosis* heat shock protein 65. MO-8, MO-9, MO-10, MO-26 and MO-29 were found to be highly related to each other and to show some homology to *M. tuberculosis* dihydrolipamide succinyltransferase. MO-28, MO-31 and MO-32 were found to be identical and to show some homology to a previously identified *M. tuberculosis* protein. MO-33 was found to show some homology to a previously identified 14 kDa *M. tuberculosis* heat shock protein.

Further studies using the above protocol resulted in the isolation of an additional four clones, hereinafter referred to as MO-12, MO-13, MO-19 and MO-30. The determined 5' cDNA sequences for these clones are provided in SEQ ID NO: 295-298, respectively, with the corresponding predicted protein sequences being provided in SEQ ID NO: 299-302, respectively. Comparison of these sequences with those in the gene bank as described above revealed no significant homologies to MO-39. MO-12, MO-13 and MO-19 were found to show some homologies to unknown sequences previously isolated from *M. tuberculosis*.

EXAMPLE 6

ISOLATION OF DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

ON September 1, 1998, the following

This example illustrates isolation of DNA sequences encoding *M. tuberculosis* antigens by screening of a novel expression library with sera from *M. tuberculosis*-infected patients that were shown to be unreactive with a panel of the recombinant *M. tuberculosis* antigens TbRa11, TbRa3, Tb38-1, TbH4, TbF and 38 kD.

Genomic DNA from *M. tuberculosis* Erdman strain was randomly sheared to an average size of 2 kb, and blunt ended with Klenow polymerase, followed by the addition of EcoRI adaptors. The insert was subsequently ligated into the Screen phage vector (Novagen, Madison, WI) and packaged *in vitro* using the PhageMaker extract (Novagen). The resulting library was screened with sera from several *M. tuberculosis* donors that had been shown to be negative on a panel of previously identified *M. tuberculosis* antigens as described above in Example 3B.

A total of 22 different clones were isolated. By comparison, screening of the λ Zap library described above using the same sera did not result in any positive hits. One of the clones was found to represent TbRa11, described above. The determined 5' cDNA sequences for 19 of the remaining 21 clones (hereinafter referred to as Erdsn1, Erdsn2, Erdsn4-Erdsn10, Erdsn12-18, Erdsn21-Erdsn23 and Erdsn25) are provided in SEQ ID NO: 303-322, respectively, with the determined 3' cDNA sequences for Erdsn1, Erdsn2, Erdsn4, Erdsn5, Erdsn7-Erdsn10, Erdsn12-18, Erdsn21-Erdsn23 and Erdsn25 being provided in SEQ ID NO: 323-341, respectively. The complete cDNA insert sequence for the clone Erdsn24 is provided in SEQ ID NO: 342. Comparison of the determined cDNA sequences with those in the gene bank revealed no significant homologies to the sequences provided in SEQ ID NO: 309, 316, 318-320, 322, 324, 328, 329, 333, 335, 337, 339 and 341. The sequences of SEQ ID NO: 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342 were found to show some homology to unknown sequences previously identified in *M. tuberculosis*.

EXAMPLE 7

ISOLATION OF SOLUBLE *M. TUBERCULOSIS* ANTIGENS

1. Preparation of Soluble Antigen

This example illustrates the use of mass spectrometry to identify soluble *M. tuberculosis* antigens.

In a first approach, *M. tuberculosis* culture filtrate was screened by Western analysis using serum from a tuberculosis-infected individual. The reactive bands were excised from a silver stained gel and the amino acid sequences determined by mass spectrometry. The determined amino acid sequence for one of the isolated antigens is provided in SEQ ID NO: 343. Comparison of this sequence with those in the gene bank revealed homology to the 85b precursor antigen previously identified in *M. tuberculosis*.

In a second approach, the high molecular weight region of *M. tuberculosis* culture supernatant was studied. This area may contain immunodominant antigens which may be useful in the diagnosis of *M. tuberculosis* infection. Two known monoclonal antibodies, IT42 and IT57 (available from the Center for Disease Control, Atlanta, GA), show reactivity by Western analysis to antigens in this vicinity, although the identity of the antigens remains unknown. In addition, unknown high-molecular weight proteins have been described as containing a surrogate marker for *M. tuberculosis* infection in HIV-positive individuals (*Jni. Infect. Dis.*, 176:133-143, 1997). To determine the identity of these antigens, two-dimensional gel electrophoresis and two-dimensional Western analysis were performed using the antibodies IT57 and IT42. Five protein spots in the high molecular weight region were identified, individually excised, enzymatically digested and subjected to mass spectrometric analysis.

The determined amino acid sequences for three of these spots (referred to as spots 1, 2 and 4) are provided in SEQ ID NO: 344, 345-346 and 347, respectively. Comparison of these sequences with those in the gene bank revealed that spot 1 is the previously identified *PcK-1*, a phosphoenolpyruvate kinase. The two sequences isolated from spot 2 were determined to be from two DNAs, previously identified in *M. tuberculosis* as heat shock proteins. Spot 4 was determined to be the previously identified *M. tuberculosis* protein *Kat G*. To the best of the inventors' knowledge, neither *PcK-1* nor the two DNAs have previously been shown to have utility in the diagnosis of *M. tuberculosis* infection.

USE OF REPRESENTATIVE ANTIGENS FOR DIAGNOSIS OF TUBERCULOSIS

This example illustrates the effectiveness of several representative polypeptides in skin tests for the diagnosis of *M. tuberculosis* infection.

Individuals were injected intradermally with 100 μ l of either PBS or PBS plus Tween 20TM containing either 0.1 μ g of protein (for TbH-9 and TbRa35) or 1.0 μ g of protein (for TbRa38-1). Induration was measured between 5-7 days after injection, with a response of 5 mm or greater being considered positive. Of the 20 individuals tested, 2 were PPD negative and 18 were PPD positive. Of the PPD positive individuals, 3 had active tuberculosis, 3 had been previously infected with tuberculosis and 9 were healthy. In a second study, 13 PPD positive individuals were tested with 0.1 μ g TbRa11 in either PBS or PBS plus Tween 20TM as described above. The results of both studies are shown in Table 10.

TABLE 10
RESULTS OF DIH TESTING WITH REPRESENTATIVE ANTIGENS

	TbH-9 Pos/Total	Tb38-1 Pos/Total	TbRa35 Pos/Total	Cumulative Pos/Total	TbRa11 Pos/Total
PPD negative	0/2	0/2	0/2	0/2	
PPD positive					
healthy	5/9	4/9	4/9	6/9	1/4
prior TB	3/5	2/5	2/5	4/5	3/5
active	3/4	3/4	0/4	4/4	1/4
TOTAL	11/18	9/18	6/18	14/18	5/13

EXAMPLE 9

SYNTHESIS OF SYNTHETIC POLYPEPTIDES

hexafluorophosphate) activation. A Gly-Cys-Gly sequence may be attached to the amino terminus of the peptide to provide a method of conjugation or labeling of the peptide. Cleavage of the peptides from the solid support may be carried out using the following cleavage mixture: trifluoroacetic acid:ethanedithiol:thioanisole:water:phenol (40:1:2:2:3). After cleaving for 2 hours, the peptides may be precipitated in cold methyl-t-butyl-ether. The peptide pellets may then be dissolved in water containing 0.1% trifluoroacetic acid (TFA) and lyophilized prior to purification by C18 reverse phase HPLC. A gradient of 0%-60% acetonitrile (containing 0.1% TFA) in water (containing 0.1% TFA) may be used to elute the peptides. Following lyophilization of the pure fractions, the peptides may be characterized using electrospray mass spectrometry and by amino acid analysis.

EXAMPLE 10

PREPARATION AND CHARACTERIZATION OF *M. TUBERCULOSIS* FUSION PROTEINS

A fusion protein containing TbRa3, the 38 kD antigen and Tb38-1 was prepared as follows.

Each of the DNA constructs TbRa3, 38 kD and Tb38-1 were modified by PCR in order to facilitate their fusion and the subsequent expression of the fusion protein TbRa3-38 kD-Tb38-1. TbRa3, 38 kD and Tb38-1 DNA was used to perform PCR using the primers PDM-64 and PDM-65 (SEQ ID NO: 146 and 147), PDM-57 and PDM-58 (SEQ ID NO: 148 and 149), and PDM-69 and PDM-60 (SEQ ID NO: 150 and 151), respectively. In each case, the DNA amplification was performed using 10 μ l 10X Pfu buffer, 2 μ l 10 mM dNTPs, 2 μ l each of the PCR primers at 10 μ M concentration, 81.5 μ l water, 1.5 μ l Pfu DNA polymerase (Stratagene, La Jolla, CA) and 1 μ l DNA at either 70 ng μ l (for TbRa3) or 50 ng μ l (for 38 kD and Tb38-1). For TbRa3, denaturation at 94°C was performed for 2 min, followed by 40 cycles of 96°C for 15 sec and 72°C for 1 min, and lastly by 72°C for 4 min. For 38 kD, denaturation at 96°C was performed for 2 min, followed by 40 cycles of 96°C for 30 sec, 68°C for 15 sec and 72°C for 3 min, and finally by 72°C for 4 min. For Tb38-1 denaturation at

1.5 min. 30 cycles of 96°C for 15 sec, 64°C for 15 sec and 72°C for 1.5, and finally by 72°C for 4 min.

The TbRa3 PCR fragment was digested with NdeI and EcoRI and cloned directly into pT7⁺L2 IL 1 vector using NdeI and EcoRI sites. The 38 kD PCR fragment was digested with Sse8387I, treated with T4 DNA polymerase to make blunt ends and then digested with EcoRI for direct cloning into the pT7⁺L2Ra3-1 vector which was digested with StuI and EcoRI. The 38-1 PCR fragment was digested with Eco47III and EcoRI and directly subcloned into pT7⁺L2Ra3/38kD-17 digested with the same enzymes. The whole fusion was then transferred to pET28b - using NdeI and EcoRI sites. The fusion construct was confirmed by DNA sequencing.

The expression construct was transformed into BLR pLys S *E. coli* (Novagen, Madison, WI) and grown overnight in LB broth with kanamycin (30 µg/ml) and chloramphenicol (34 µg/ml). This culture (12 ml) was used to inoculate 500 ml 2XYT with the same antibiotics and the culture was induced with IPTG at an OD₅₆₀ of 0.44 to a final concentration of 1.2 mM. Four hours post-induction, the bacteria were harvested and sonicated in 20 mM Tris (8.0), 100 mM NaCl, 0.1% DOC, 20 µg/ml Leupeptin, 20 mM PMSF followed by centrifugation at 26,000 X g. The resulting pellet was resuspended in 8 M urea, 20 mM Tris (8.0), 100 mM NaCl and bound to Pro-bond nickel resin (Invitrogen, Carlsbad, CA). The column was washed several times with the above buffer then eluted with an imidazole gradient (50 mM, 100 mM, 500 mM imidazole was added to 8 M urea, 20 mM Tris (8.0), 100 mM NaCl). The eluates containing the protein of interest were then dialyzed against 10 mM Tris (8.0).

The DNA and amino acid sequences for the resulting fusion protein hereinafter referred to as TbRa3-38 kD-Tb38-1 are provided in SEQ ID NO: 152 and 153, respectively.

A fusion protein containing the two antigens TbH9 and Tb38-1 (hereinafter referred to as TbH9-Tb38-1) without a hinge sequence, was prepared using a similar procedure to that described above. The DNA sequence for the TbH9-Tb38-1 fusion protein is provided in SEQ ID NO: 154.

The ability of the fusion protein TbH9-Tb38-1 to induce T cell proliferation and IFN- γ production in PBMC preparations was examined using the protocol described above in Example 1. PBMC from three donors were employed: one who had been previously shown to respond to TbH9 but not Tb38-1 (donor 131); one who had been shown to respond to Tb38-1 but not TbH9 (donor 184); and one who had been shown to respond to both antigens (donor 201). The results of these studies (Figs. 5-7, respectively) demonstrate the functional activity of both the antigens in the fusion protein.

A fusion protein containing TbRa3, the antigen 38kD, Tb38-1 and DPEP was prepared as follows.

Each of the DNA constructs TbRa3, 38 kD and Tb38-1 were modified by PCR and cloned into vectors essentially as described above, with the primers PDM-69 (SEQ ID NO:150 and PDM-83 (SEQ ID NO: 205) being used for amplification of the Tb38-1A fragment. Tb38-1A differs from Tb38-1 by a DraI site at the 3' end of the coding region that keeps the final amino acid intact while creating a blunt restriction site that is in frame. The TbRa3 38kD Tb38-1A fusion was then transferred to pET28b using NdeI and EcoRI sites.

DPEP DNA was used to perform PCR using the primers PDM-84 and PDM-85 (SEQ ID NO: 206 and 207, respectively) and 1 μ l DNA at 50 ng/ μ l. Denaturation at 94 °C was performed for 2 min, followed by 10 cycles of 96 °C for 15 sec, 68 °C for 15 sec and 72 °C for 1.5 min; 30 cycles of 96 °C for 15 sec, 64 °C for 15 sec and 72 °C for 1.5 min; and finally by 72 °C for 4 min. The DPEP PCR fragment was digested with EcoRI and Eco721 and clones directly into the pET28Ra3/38kD/38-1A construct which was digested with DraI and EcoRI. The fusion construct was confirmed to be correct by DNA sequencing. Recombinant protein was prepared as described above. The DNA and amino acid sequences for the resulting fusion protein hereinafter referred to as TbF-21 are provided in SEQ ID NO: 208 and 209, respectively.

The reactivity of the fusion protein TbF-21 with sera from *M. tuberculosis*-infected patients was examined by ELISA using the protocol described above. The results of these studies (Table 11) demonstrate that all four antigens function independently in the fusion protein.

TABLE 11
REACTIVITY OF TbF-2 FUSION RECOMBINANT WITH TB AND NORMAL SERA

Serum ID	Status	TbF OD450	Status	TbF-2 OD450	Status	ELISA Reactivity			
						38 kD	TbRa5	Tb38-1	DPEP
B931-40	TB	0.57	-	0.321	-	-	-	-	-
B931-41	TB	0.601	-	0.396	-	-	-	-	-
B931-109	TB	0.494	-	0.404	-	-	-	-	-
B931-132	TB	1.502	-	1.292	-	-	-	-	-
5004	TB	1.806	-	1.666	-	-	-	-	-
15004	TB	2.862	-	2.468	-	-	-	-	-
39004	TB	2.443	-	1.722	-	-	-	-	-
68004	TB	2.871	-	2.575	-	-	-	-	-
09004	TB	0.691	-	0.971	-	-	-	-	-
107004	TB	0.875	-	0.732	-	-	-	-	-
92004	TB	1.632	-	1.394	-	-	-	-	-
97004	TB	1.491	-	1.979	-	-	-	-	-
118004	TB	3.182	-	3.045	-	-	-	-	-
173004	TB	3.644	-	3.578	-	-	-	-	-
175004	TB	3.332	-	2.916	-	-	-	-	-
274004	TB	3.696	-	3.716	-	-	-	-	-
276004	TB	3.243	-	2.56	-	-	-	-	-
282004	TB	1.249	-	1.234	-	-	-	-	-
289004	TB	1.573	-	1.17	-	-	-	-	-
308004	TB	3.708	-	3.355	-	-	-	-	-
314004	TB	1.665	-	1.399	-	-	-	-	-
317004	TB	1.163	-	0.92	-	-	-	-	-
312004	TB	1.709	-	1.455	-	-	-	-	-
380004	TB	0.238	-	0.461	-	-	-	-	-
451004	TB	0.18	-	0.2	-	-	-	-	-
478004	TB	0.188	-	0.469	-	-	-	-	-
410004	TB	0.384	-	2.392	-	-	-	-	-
411004	TB	0.306	-	0.874	-	-	-	-	-
421004	TB	0.357	-	4.56	-	-	-	-	-
528004	TB	0.047	-	0.196	-	-	-	-	-
A6-87	Normal	0.094	-	0.063	-	-	-	-	-
A6-88	Normal	0.214	-	1.19	-	-	-	-	-
A6-89	Normal	0.248	-	0.125	-	-	-	-	-
A6-90	Normal	0.179	-	0.296	-	-	-	-	-
A6-91	Normal	0.133	-	0.151	-	-	-	-	-
A6-92	Normal	0.064	-	0.097	-	-	-	-	-
A6-93	Normal	0.072	-	0.098	-	-	-	-	-
A6-94	Normal	0.072	-	0.064	-	-	-	-	-
A6-95	Normal	0.125	-	0.189	-	-	-	-	-
A6-96	Normal	0.121	-	0.12	-	-	-	-	-

A fusion protein containing TbRa3, the antigen 38kD, Tb38-1 and TbH4 was prepared as follows.

Genomic *M. tuberculosis* DNA was used to PCR full-length TbH4 (FL TbH4) with the primers PDM-157 and PDM-160 (SEQ ID NO: 348 and 349, respectively) and 2 μ l DNA at 100 ng/ μ l. Denaturation at 96 °C was performed for 2 min, followed by 40 cycles of 96 °C for 30 sec, 61 °C for 20 sec and 72 °C for 5 min; and finally by annealing at 72 °C for 10 min. The FL TbH4 PCR fragment was digested with EcoRI and Sca I (New England Biolabs.) and cloned directly into the pET28Ra3/38kD/38-1A construct described above which was digested with DraI and EcoRI. The fusion construct was confirmed to be correct by DNA sequencing. Recombinant protein was prepared as described above. The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbF-6) are provided in SEQ ID NO: 350 and 351, respectively.

A fusion protein containing the antigen 38kD and DPEP separated by a linker was prepared as follows.

38 kD DNA was used to perform PCR using the primers PDM-176 and PDM-175 (SEQ ID NO: 352 and 353, respectively), and 1 μ l PET28Ra3/38kD/38-1/Ra2A-12 DNA at 110 ng/ μ l. Denaturation at 96 °C was performed for 2 min, followed by 40 cycles of 96 °C for 30 sec, 71 °C for 15 sec and 72 °C for 5 min and 40 sec; and finally by annealing at 72 °C for 4 min. The two sets of primers PDM-171, PDM-172, and PDM-173, PDM-174 were annealed by heating to 95 °C for 2 min and then ramping down to 25 °C slowly at 0.1 °C/sec. DPEP DNA was used to perform PCR as described above. The 38 kD fragment was digested with Eco RI (New England Biolabs) and cloned into a modified pET7AL2 vector which was cut with Eco 72 I (Promega) and Eco RI. The modified pET7AL2 construct was designed to have a MGHHHHHH amino acid coding region in frame just 5' of the Eco 72 I site. The construct was digested with Kpn 2I (Gibco, BRL) and Pst I (New England Biolabs) and the annealed sets of phosphorylated primers (PDM-171, PDM-172 and PDM-173, PDM-174) were cloned in. The DPEP PCR fragment was digested with Eco RI and Eco 72 I and cloned into this second construct which was digested with Eco 47 III (New England Biolabs) and Eco RI. Location of the linker is indicated in the following diagram.

modified pET28 vector. The fusion construct was confirmed to be correct by DNA sequencing.

Recombinant protein was prepared essentially as described above. The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbF-8) are provided in SEQ ID NO: 354 and 355, respectively.

One of skill in the art will appreciate that the order of the individual antigens within the fusion protein may be changed and that comparable activity would be expected provided each of the epitopes is still functionally available. In addition, truncated forms of the proteins containing active epitopes may be used in the construction of fusion proteins.

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for the purpose of illustration, various modifications may be made without deviating from the spirit and scope of the invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANTS: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald

(ii) TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS

(iii) NUMBER OF SEQUENCES: 355

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: SEED and BERRY LLP
(B) STREET: 6300 Columbia Center, 701 Fifth Avenue
(C) CITY: Seattle
(D) STATE: Washington
(E) COUNTRY: USA
(F) ZIP: 98104-7092

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.3, Version #1.30

vi. CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE: 05-MAY-1998
(C) CLASSIFICATION:

viii. ATTORNEY/AGENT INFORMATION:

A. NAME: MARK DAVID E.
B. REGISTRATION NUMBER: 41,192
C. REFERENCE DOCUMENT NUMBER: 02/011,411/9

(ix) TELECOMMUNICATION INFORMATION

(A) TELEPHONE: 206 522-4900
(B) TELEFAX: 206 522-6231

x. INFORMATION FOR SEQ ID NO. 1

a. SEQUENCE CHARACTERISTICS:

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO 1:

CGAGGCACCG	GTAGTTTGAA	CCAAACGCAC	AATCGACGGG	CRAACGAACG	GAAGAACACA	60
ACCATGAAGA	TGGTGAAATC	GATCGCCGCA	GGTCTGACCG	CCGCGGCTGC	AATCGGCGCC	120
GCTGCGGCCG	GTGTGACTTC	GATCATGGCT	GGCGGCCCGG	TCGTATACCA	GATGCAGCCG	180
GTCTCTTCG	GCGCGCCACT	GCCGTGGGAC	CCGGCATCCG	CCCCTGACGT	CCCGACCGCC	240
GCCCAGTTGA	CCAGCCTGCT	CAACAGCCTC	GCCGATCCCA	ACGTGTCGTT	TGCGAACAAG	300
GGCAGTCTGG	TCGAGGGCGG	CATCGGGGGC	ACCGAGGCGC	GCATCGCCGA	CCACAAGCTG	360
AAGAAGGCCG	CCGAGCACCG	GGATCTGCCG	CTGTCTTCA	GCGTGACGAA	CATCCAGCCG	420
GCGGCCGCCG	GTTCTGGCCAC	CGCCGACGTT	TCCGTCTCGG	GTCCGAAGCT	CTCGTCGCCG	480
GTCAACGAGA	ACGTCACGTT	CGTGAATCAA	GGCGGCTGGA	TGCTGTCACG	CGCATCGGCG	540
ATGGAGTTGC	TGCAGGCCGC	AGGGNAACTG	ATTGGCGGGC	CGGNTTCAGC	CCGCTGTTCA	600
GCTACGCCGC	CCGCCTGGTG	ACGCGTCCAT	GTCAACACT	CGCGCGTGTA	GCACCGTGCG	660
GTNTGCCGAG	GGNCSCACGC	ACCGCCCGGT	GCAAGCCGTC	CTCGAGATAG	GTGGTGNCTC	720
GNCAACAGNG	ANCACCCCTN	NNTCGNCINT	TCTCGNTGNT	GNATGA		766

(2) INFORMATION FOR SEQ ID NO:1:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 762 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ATGCATCACC	ATCACCATCA	CGATGAAGTC	ACGCTAGAGA	CGACCTCCGT	CTTCCGCGCA	60
GACTTCCTCA	GCGAGCTGGA	CGCTCTCTGG	CAAGCGGGTA	CGGAGAGCGC	GGTCTCCGGG	120
GTGGAAGGGC	TCCCGCCCGG	CTCGGCGTTG	GTGGTAGTCA	AACGAGGCCC	CAACGCCGGG	180
TCCCGGTTCC	TACTCGACCA	AGGCATCAGC	TCGGGTGGTC	GGCATCCCGA	CAGCGACATA	240
TTCTCTGACG	ACGTGACTGT	GAGCGCTCGC	CATGCTGAAT	TCCCGTTGGA	AAACAACGAA	300
TTCAATGTCC	TGGATGTCCG	GAGTGTGAAT	GCGACCTAGC	TCAACCGCGA	GCCCCGTGGT	360
TGGGCGGTGC	TGGCGAACGG	CGACGAGGTC	CAGATCGGCA	AJCTCCGCTT	GGTGTCTTTC	420
ACTGGACCCA	AGCAAGCGCA	GGATGACGGG	AGTACGCGGC	GCCCGTGAGC	GCACCCGATA	480
GCCCCCGGCT	GCCCCGGATG	TCGATCGGGG	CGGTCTCTCC	ACTGCTACG	ACCGGATTTT	540
CCCTGATGTC	CACCATCTCC	AAGATTGCGT	TCTTGGGAGC	CTTGAGGGTC	NGGGTGACCC	600
CCCTGCGGGC	CTCATTGCGG	GCTNTCGGCT	GCTTTCAGCC	CTTACCTACT	GCTNCCCGCT	660
TTTGNAAATTC	TTTCTTCTCT	GCTTNNAAAT	TTTCTTCTCT	TTTCTTCTCT	TTTCTTCTCT	720
TTTCTTCTCT	TTTCTTCTCT	TTTCTTCTCT	TTTCTTCTCT	TTTCTTCTCT	TTTCTTCTCT	762

(2) INFORMATION FOR SEQ ID NO:2:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 813 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:3:

(A) LENGTH: 447 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

CGGTATGAAC	ACGGCCGGCT	CGGATAACTT	CCAGCTGTCC	CAGGGTGGGC	ACGGATTCCG	60
CATTCCGATC	GGGCAGGCGA	TGGCGATCCG	GGGCCAGATC	CGATCGGGTG	GGGGGTCAAC	120
CACCGTTTCT	ATCGGGCCTA	CCCGCTTCCT	CGGCTTGGGT	GTGTGCGACA	ACAACGGCAA	180
CGGCGCACGA	GTCCAACCGG	TGTCGGGAG	CGCTCCGGCG	GCAAGTCTCG	GCATCTCCAC	240
CGGCGACCTG	ATCACCGCGG	TGACGGCGG	TCCGATCAAC	TGGGCCACCG	CGATGGCGGA	300
CGCGCTTAAC	GGGCATGATC	CGGTGAGGT	CATCTCGGTG	AACTGGCAAA	CGAAGTCGGG	360
TGACCGCGT	ACAGTGAACG	TGATATGGC	CGAGGGACCC	CGGGCTGAT	TTCGTGCGGG	420
ATACGACCGG	CCGGCCGGCG	AATTGGA				440

A) LENGTH: 604 base pairs
B) TYPE: nucleic acid
C) STRANDEDNESS: single
D) TOPOLOGY: linear

[illegible]

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 633 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```

TTGCANGTCG AACCACCTCA CTAAAGGGAA CAAAAGCTNG AGCTCCACCG CCGTGGCGGC      60
CGCTCTAGAA CTAGTGKATM YYYCKGGCTG CAGSAATYCG GYACGAGCAT TAGGACAGTC      120
TAACGGTCTCT GTTACGGTGA TCGAATGACC GACGACATCC TGCTGATCGA CACCGACGAA      180
CGGGTGCGAA CCTCACCT CAACCGGCCG CAGTCCCGYA ACACGCTCTC GCGCGCGCTA      240
CGGGATCGGT TTTTCGCGGY GTTGGYCGAC GCCGAGGYCG ACACGACAT CGACGTCTGC      300
ATCCTCACCG GYCCGATCC GGTGTTCTGC GCCGACTGG ACCTCAAGGT AGCTGGCCCG      360
GCAGACCGCG CTGCGGACA TCTACCCCG GTGGGCGGCC ATGACCAAGC CCGTGATCGG      420
CGCGATCAAC GCGCGCGCG TCACCGCGCG GCTCGAAGTG GCGGTGTACT GCGACATCCT      480
GATCGCTTCC GAGGACGCG GCTTCGNCGA CACCGACGCG TGGGTGGCGG TGCTGCGCAC      540
TGGGGGACTC AGTGTGTGCT TCGCGCAAAA GGTGGGCGAT GGNATGGCGG GGTGGATGAG      600
CGTGACCGCG GACTACCTGT CGTGACCGCA CGC

```

(x1) INFORMATION FOR SEQ ID NO 7:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1362 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```

CGACGACGAC GCGGCGCGAG AGCGGGCGCG AACGGCGATC GACGCGCGCC TGGCGAGAGT      60
TGGCACCACT TGGAGGGGAG TCGAATCATG AAATTGTGCA ACCATATTGA GCGCGTGGCG      120
TGGCGCGGAG GCGGCGCGCG GGTGGCGGAG GTTATGCGCG AGCGCGCGCG CGAGTTCGCG      180
TGGGTGCGCG AGTGGCTGCG CATGCTGTGCT TGGACGAGG GATGCTCTAC GCGCGGCTGG      240
GCGACGTTGC GCGAGAGACT GCTGGTGGGC GAGGTGCGCG GTGGCGCGCA GGAAGCGCTG      300
GCGCGCGCGG TGGCGCGGAG GCTGGCGTGC TGGTGGTGC TCGACGCGCA CACCACCATG      360
GTGTACGCGG CAGGCGCAAA GACACCGTGC GCGGCGATCT TGGCGCGGAC AGCACCTGCC      420
GCGGGTGACC CGAAGCGCGT GATGTGGCG TGGCGCGGAG GAACCGGGAC ACCGGCGGGA      480
TGGCGCGGAC GTTGGCGCGG GATGTGGCG TGGCGGAGAG TGGCGAGCGG GGTGCAATTC      540
GACTTCATCG GACTGCTGGT GCTGCTGCTG TGGAGGAAA GTTGGCTGCG GGGGGCGCGG      600
GCGCGCGCAAG AGCTGATGCG GCGCGCGGCT GCACTGCTGT TGGCGCGGAA GGTGCGCGCG      660
GAGCATGCGG GCGGCGGCTG CACCGCGCGG TCGAGGCTG GAAAGCTGCG CGACGATCTG      720
GATGGGCAA GAGGCTGCGA GCGGCTGCGA CCGGCTTCT GCGGCTGCGA CACCGCTG      780
GACACCGCGG GCGGCTGCGG GCGGCGACT GCTAGGCTG TCGGCGGCTG GGTGGGCTG      840
TGGGCGGCGG AGGCAATGCG GATGAGAGT GCTGGAGCA ACAGGACAC CCGCGAGCTG      900
GCGGCGGAGG TCGAGCGCGG CACCGCTGCT GCGGTGCTGA GCGGCTGCG CCGGATCAG      960
GTGAGCGAGG AGGAGCTGCT GCGGCGCGCA TGGGTGCTG ACAGGAGTGC GCGGCTGCTT      1020
TGGCGCTGCT GCGGCGCGCG GTTACCGCG GCGGCGCGCA TCGGCGGCTG GATGCGCGCG      1080
GCGGCGGAGG GCGAGGCTG GCGGCGGAAA GCGGCTGCTG GAGTGTGCTG GCGGCTGCTG      1140
GAGGCTGCTG GCGGCTGCTG GCGGCTGCTG GCGGCTGCTG GCGGCTGCTG GCGGCTGCTG      1200
GAGGCTGCTG GCGGCTGCTG GCGGCTGCTG GCGGCTGCTG GCGGCTGCTG GCGGCTGCTG

```


11. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1458 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```

GCGACGACCC CGATATGCCG GGCACCGTAG CGAAAGCCGT CGCCGACGCA CTCGGGCGCG      60
GTATCGCTCC CGTTGAGGAC ATTCAGGACT GCGTGGAGGC CCGGCTGGGG GAAGCCCGTC      120
TGGATGACGT GGGCCGTGTT TACATCATCT ACCGGCAGCG GCGCGCCGAG CTGCGGACGG      180
CTAAGGCCCTT GCTCGGCGTG CGGGACGAGT TAAAGCTGAG CTTGGCGGCC GTGACGGTAC      240
TGC3CGAGCG CTATCTGCTG CACGACGAGC AGGGCCGGCC GGGCGAGTCG ACCGGCGAGC      300
TGATGGACCG ATCGGCGCGC TGTGTGCGCG GGGCGAGGA CCAGTATGAG CCGGGCTCGT      360
CGAGGCGCGT GGGCGAGCGG TTCGCCACGC TATTACGCAA CCTGGAATTC CTGCCGAATT      420
CGCCACCGTT GATGAACCTT GGCACCGAGT TGGAGCTGCT GCGCGGCTGT TTTGTTCTGC      480
CGATTGAGGA TTGCTGCAA TCGATCTTTG GGCAGCTGGG ACAGGCGCGC GAGCTGCGAC      540
GGGCTGAGGG CGGCACCGGA TATGCGTTCA GGCACCTGGC ACCCGCGCGG GATCGGGTGG      600
GCTCCACGGG CGGCACCGCC AGCGGACCGG TGTGCTTTCT ACCGCTGTAT GACAGTGGCG      660
CGGGTGTGGT CTCCATGGGC GGTGCGCGGC GTGGCGCGTG TATGGCTGTG CTTGATGTGT      720
CGACCCCGGA TATCTGTGAT TTGCTCACCG CCAAGGCCGA ATCCCCCAGC GAGCTCCCGC      780
ATTTCACACT ATCGGTGGT GTGACCGACG CTTTCCTGCG GCGCGTCGAA CGCAACGGCG      840
TACACCGGCT GGTCAATCGG CGAACCGGCA AGATCGTCGG GCGGATGCCC GCGCGCGAGC      900
TGTTCGACCG CATCTGCAA GCGCGCGCAG CCGGTGGCGA TCGCGGGCTG GTGTTCTCG      960
ACACGATCAA TAGGGCAAAC CCGGTGCGCG GGAGAGGCGG CATCGAGCGC ACCAAGCCGT      1020
CGCGGGAGGT CCGACTGCTG CTTACGACT GATCTAATCT CGGCTCGATC AACCTCGCCC      1080
CGATGCTCGC CGACGCTCGG GTGACTGGG ACCGGCTCGA GGAGGTGCGC GGTGTGGCGG      1140
TCCGGTTGCT TGATGACGTC ATCGATGTCA GCGGCTACCC GTTCCGCGAA CTGGGTGAGG      1200
CGCGCGCGCG CACCCGCAAG ATCGGGTGG GACTCATGGG TTTGGCGGAA CTGCTTGGCG      1260
CAGTGGGTAT TCGGTACGAC AGTGAAGAA CCGTGGCGTT AGCGACCGGG CTCATGCGTC      1320
GCATACAGCA GGTGGCGCAC ACGGCATCGG CGAGGCTGGC CGAAGAGGCG GCGCGATTCC      1380
CGCGCTTCAG GCATACCGCG TTGCGCGCTT CGCGCGCGAG TGTACGCA CAGGTCACT      1440
CGGTGCTGCT GACGGGCA

```

2. INFORMATION FOR SEQ ID NO:4

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 962 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

11. SEQUENCE DESCRIPTION: SEQ ID NO:4

```

ACGCTGTAAT CTTGCTGAT CTGGAACCGT TTGGGCTGCT ACCTACGAG ATCTACTGCG      60
CGCCGAGGGG GTTGGGCTGT GGCATCGGCG TGTGCTAGT GGGGATCGCG GTGGGATCG      120
TCATCGGCTT CTTGACAGG AGCCCGGCTG CGAAACCGGT TADGGCGGAC AAGTGGGCTT      180
CGCCCGAGAG CCACTCGGCG TCGCGGCGCA TTAAGGAGG TCGGCGGCTT GGGCAAGCG      240
AAGCTAACCG CGCTGCTG TCGGCGGCTT TTAAGGAGG TCAAGAGAGT AGGTGAGCG      300
CGCGGCTGCA GCGGCTGCT TCGGCGGCTT TTAAGGAGG TCAAGAGAGT AGGTGAGCG      360

```

GGACCGGGAT GGGATCGGGG CCGCGCTGCC CATTCGCCCG GCGGGCGATG GGGCGGGGCA	660
CCTACAAATCT CGTGGTACAA CTGGGCAATG TCGGCTCGCT GCGGTTCCG TTCATCCTGA	720
ATCAGCCCGC GCGCGCGCGG GGGCGGTAC CCGCTCCGGG TCCAGCGCAG GCGCCTCCGC	780
CGGAGTCTCC CCGCGAAGGC GGATAATTAT TGATCGCTGA TGGTCGATTC CGCCAGCTGT	840
GACAACCCCT CGCCTCGTGC CG	862

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 622 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TTGATCAGCA CCGGCAAGGC GTCACATGCC TCGCTGGGTG TCGAGGTGAC CAATGACAAA	60
AGACGCGCGG GCGCCAAGAT CGTCCAAGTA GTGGCGCGGTG GTGCTGCCGC GAACGCTGGA	120
GTGGCGAAGG CGCTCGTTGT CACCAAGGTC GACGACCGCC CGATCAACAG CCGCGACCGG	180
TTGGTTGCCG CGCTGCGGTG CAAAGCGCCG CGCGCCACGG TGGCGCTAAC CTTTCAGGAT	240
CGCTCGGGCG GTAGCCGCAC AGTGCAAGTC ACCCTCGGCA AGGCGGAGCA GTGATGAAGG	300
TGCGCGSCA GTGTTCAAAG CTCGGATATA CGGTGGCACC CATGGAACAG CGTGGCGAGT	360
TGGTGGTTGG CCGGGCACTT GTCGTCTGTC TTGACGATCG CACGGCGCAC GCGCATGAAG	420
ACCACAGCGG GCGGCTTGTG ACCGAGCTGC TCACCGAGGC CCGGTTTGTG GTCGACGGCG	480
TGGTGGCGGT GTCGGCGCAC GAGGTGAGAG TCCGAAATGC GCTGAACACA GCGGTGATCG	540
GCGGGGTGGA CGTGGTGGTG TCGGTGGCGG GGACCGNGT GACGCTCGCG GATGTCAACC	600
CGGAAGCCAC CCGGACATT CT	622

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- A. LENGTH: 1200 base pairs
 B. TYPE: nucleic acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AGCGAGCGG TAAGCGTGTG GCGCGCGCGC ACAGTGTGTG TGACAGGATG GGGCGGTGCG	60
AGCAGAGCTT CGTCTGAGG AGCAGCGCGA AGCTGTGGGT CGGTGCATG CCGCGCGAAG	120
AAAGAGCTCC AGTCCAGCGG ATGACCGGCA AAGAAAATG CCATCGAGCA GTTCCTGTAT	180
AGGTACCTGC GATCGTGGCG GCGTACAGG TTGGATACA AGCGCAACGG GTCCGCTGCG	240
AGGTGACCC AGTTTGTCAA AAAGAAAAC GATTCGCGG GTTCGGATGT CCGCTTGAAT	300
AGGTGACCC GTCAAGCTGA AGGTGCGCG GAGCGGTGCG GTTCGCGCGC ATGGGACGTC	360
AGGTGCTGT TCGGCGCGAT AGCATCAGT TACAAATATCA AGGCGGTGAG CACGCTGAAT	420
AGGTGACGAC CCACTACCGT CAAGATTTTG AAGCGGACCA TCACCGTGTG GAATGATGCA	480
AGATCGAAG CCGTCAACTG CGGACCGGAC GTCGCGCGAA CACCGATTAG CGTTATCTTC	540
AGAGGAGACA AGTCCGCTAC GTCGGACAGC TTCCAGAAAT ACCTCGACCG TGTATCCAAC	600
AGGGGTGGG GCAAAGCGCG TAGGAAAACG TGCAGCGGGG GGTTCGCGCT CCGCGCGCAG	660
AGAAAGAGG GAAAGCTGGG GTATGCGAA AGTACGAGG GGTTCGATCA CTACAAGTAA	720
AGGTGCTGT GGTGCTGTA AGTTCGAA AGTTCGAA AGTTCGAA AGTTCGAA	780

GACCAATACG GGTCCATTCC GTTGCCCAAA TCGTTCCAAG CAAAATTGGC GGCCGCGGGT	1080
AATGCTATTT GTTGACCTAG TGAAGGGAAT TCGACGCTGA GCGATGCCGT TCCGCAGGTA	1140
GGGTGCGCAAT TTGGGGCGTA TCAGCTATTG CGGCTGCTGG GCCGAGGCGG GATGGGCGAG	1200

(2) INFORMATION FOR SEQ ID NO:12:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1155 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GCAAGCAGCT GCAGGTCGTG CTSTTCGACG AACTGGGCAT GCCGAAGACC AAACGCACCA	60
AGACCGGCTA CACCACGGAT GCCGACCGCG TGCAGTCGTT GTTCSACAAG ACCGGGCATC	120
CGTTTCTGCA ACATCTGCTG GCCCACCCTG ACCTCACCCT GCTCAAGGTC ACCGTGACCG	180
GGTGGCTGCA AGCGSTGCGC GCGGACGCGC GCATCGACAC CACGTTCAAC CAGACCATCG	240
CGCGCAGCTG CGCGCTCTCC TCGACCGAAC CCAACCTGCA GAACATCCCG ATCCGACCGC	300
ACGCGGGGCG CGCGATCCCG GACGCTTCC TGGTCGGGGA CGGTTACCGC GAGTTGATGA	360
CGGCGGACTA CAGCCAGATC GAGATGCGGA TCATGGGGCA CCTGTCCGGG GACGAGGGCC	420
TCATCGAGGC GTTCAACACC GGGGAGGACC TGTATTCTGT CGTCCGCTCC CGGGTGTTCG	480
GTGTGCCCCAT CGACGAGGTC ACCGGCGAGT TCGCGCGCGC GGTCAAGGCG ATGTCTTACG	540
AGGCGAACCA GCAGATGGAC GCGTATTTCG CCGGATTCGG CGGGGTGCGC GACTACCTGC	600
CGCGCTAGT CGAGCGGGCC CGCAAGGACG GCTACACCTC GACGGTGCTG GGCCGTCGCC	660
GCTACCTGCC CGAGCTGGAC AGCAGCAACC GTCAAGTCCG GGAGGCGGCC GAGCGGGCGG	720
CGCTGAACCG CGCGATCCAG GGCAGCGCGC CCGACATCAT CAAGGTGGCC ATGATCCAGG	780
TCGACAAAGC GGTCAACGAG GCACAGCTGG CGTCGCCCAT GCTGCTGCAG GTCCACGACG	840
AGCTGCTGTT CGAAATCCCG CCGGCTGAAC CGCAGCGCGT CGAGGCCCTG GTGCCGACA	900
AGATGGGCGG CGCTTACCGC CTCGACCTCC CGCTCGAGST CTCGGTGGGC TACGGCCGCA	960
GCTGGGACCG CGCGGCGGAC TGAGTGGCGA GCTGCTATC CGGGCGGGAA TTCGGCATT	1020
TTTGGCGCTT GAGTTCAAGC TCGCGCAAT TGGGACCGAG TTTCTCAGC GTGTACCGCT	1080
CGAGTAGCCT GTCA	1140
	1155

(2) INFORMATION FOR SEQ ID NO:13:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1171 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GAGTCTCTTC TGGTCTTTGA AGGTTTTCG CGGTGGGCAT CGGCACCGGC GTTCTCGGGT	60
TCTGCTCTTC GGTGCGCAT GTTCAACAG GTGCTCTTA AGCAGCGCGG ATGCTCTCGC	120
ATCGAAGACA CGACCCAGG TGGCAGCGCC CCGTGAACCT CGATTACCT GCTGCTCTTC	180
CGCGCTCTTA TCCGATTTC TACCTCTCC GGTGCGACCT GTGCGCTCG GATGCTCTTC	240
ATCGAAGAT CTGCGGCTTC GCGAAGCTT ATCTGCTTC AATCTGCTC CAGGCTCTTC	300
AATGCTCTTC CGAGTTACTC TCGAAGCTT CAGGCTCTTC CAGGCTCTTC CAGGCTCTTC	360

GACCATGACG	CCCCCTCCTG	GGATGGTTGG	CCAACGCCCT	CGTGCAGGCA	TGTTGGCCAT	660
CGGCGCGGTG	ACGATAGCGG	TGGTGTCCGG	CGGCATCGGC	GGCGCGGCCG	CATCCCTGGT	720
CGGGTTCAAC	CGGGCACCCG	CCGGCCCCAG	CGGCGGCCCA	GTGGCTGCCA	GCGCGGCGCC	780
AAGCATCCCC	GCAGCAAACA	TGCCGCCGGG	GTCCGGTCGA	CAGGTGGCGG	CCAAGGTGGT	840
GCCCAGTGTC	GTCTGTGG	AAACCGATCT	GGGCGGCCAG	TCCGAGGAGG	GCTCCGGCAT	900
CATTCTGTCT	GCCGAGGGGC	TGATCTTGAC	CAACAACCAC	GTGATCGCGG	CGGCCGCCAA	960
GCCTCCCCCTG	GGCAGTCCGC	CGCCGAAAAC	GACGGTAACC	TTCTCTGACG	GGCGGACCGC	1020
ACCCCTTCACG	GTGGTGGGGG	CTGACCCAC	CAGTGATATC	GCCGTCTGCC	GTGTTGAGGG	1080
CGTCTCCGGG	CTCACCCCGA	TCTCCCTGGG	TTCTCTCTCG	GACCTGAGGG	TCCGTGAGCC	1140
GGTGTGCGG	ATCGGGTCCG	CGCTCGGTTT	GGAGGGCACC	GTGACCACGG	GGATCGTCAG	1200
CSCTCTCAAC	CGTCCAGTGT	CGACGACCGG	CGAGGCCGGC	AACCAGAACA	CCGTGCTGGA	1260
CGCCATTTCAG	ACCGACGCGG	CGATCAACCC	CGGTAACCTC	GGGGGCGCGC	TGGTGAACAT	1320
GAACGCTCAA	CTCGTCCGAG	TCAACTCGGC	CATTGCCACG	CTGGGCGCGG	ACTCAGCCGA	1380
TGCCGAGAGC	GGCTCGATCG	GTCTCGGTTT	TGCGATTCCA	GTGACCCAGG	CCAAGCGCAT	1440
CGCCGACGAG	TTGATCAGCA	CGGCAAGGC	GTCCATGCCC	TCCCTGGGTG	TGCAGGTGAC	1500
CAATGACAAA	GACACCCCGG	CGCCCAAGAT	CGTCCAAGTA	GTGGCCCGGT	GTGCTGCCCG	1560
GAACCCCTGA	GTGCCGAAGG	CGCTCGTTGT	CACCAAGGTC	GACGACCGCC	CGATCAACAG	1620
CGCCGACCGG	TTGGTTGCCG	CGCTGCCGTC	CAAAGCCCGG	GGGCGCCACG	TGGCGCTAAC	1680
CTTTCAGGAT	CCCTCGGGCG	GTAGCCGCAC	AGTGCAAGTC	ACCCCTCGGC	AGGCGGAGCA	1740
GTGATGAAGG	TCCCGCCGCA	GTGTTCAAAG	C			1771

(2) INFORMATION FOR SEQ ID NO:14:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1058 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CTCCACCGCG	GTGGCGGCCG	CTCTAGAAGT	AGTGGATGCC	CGGGGCTGCA	GGAAATCGGC	90
ACGAGGATCC	GACGTCCGAG	CTTGTGGAAC	TCCCGCCCGC	CGAAGTATCG	GTCCATGCCT	120
AGCCCGCGGA	CGGCGAGCGC	CGGAATGGCG	CGAATGAGGA	GGCGGGCAAT	TTGGCGGGGC	180
CGCCCGACCG	CGAGCCCGCG	AATGGCGCGA	GTGAGGAGGC	CGCCAGTCAT	GCCACGCGTG	240
ATCCAAATCAA	CCTGCATTCC	GCCTGCCGGC	GCATTTGACA	ATCCAGGTAG	TGAGCGCAAA	300
TGAATGATGG	AAAACCGGGG	GTGACGTCCG	GTGTTCTGGT	GGTGTCTAGG	GCCTGCCTGG	360
CGTTGTGGCT	ATCAGGATGT	TCTTCGCCGA	AACCTGATGC	CGAGGAACAG	GGTGTTCGCC	420
TGAGCCCGAC	GCCCTCCGAG	CGCCCGCTCC	TGTCGAGAT	CAGGCAGTCC	CTTGATGCCA	480
MAAAGGGTT	GACGAGCTTG	TCCTAGCCCG	TCTCAACAC	CGGCAAGTCC	GACAGCTTGC	540
TGGGTATTAC	CAGTCCCGAT	GTCCACCTCC	GGGCAATCC	CTCCCGCGCA	AAGGGCGTAT	600
GCACCTACAA	CGACGAGCGG	GTGTCCCGCT	TTCCGGGTACA	AGGCGACAC	ATCTCGGTGA	660
AACTGTTCCA	CGACTGGAGT	AACTCCGGCT	CGATTTCTGA	ACTGTCAACT	TCACGCGTGC	720
TGATCCCTGC	CGCTGGGCTG	AGCCAGCTTC	TCTCGGGTGT	CACGAACCTC	CAAGCGCAAG	780
GTACCGAAGT	GATAGACCGA	ATTTCGACCA	CGAAAATCAC	CGGAGCCTCC	CCCGCGAGCT	840
GTGTCAAGAT	CGTTGATCCG	GGGCGCAAGA	GTGCAAGGCT	GGGAGCCGTC	TGGATTGCCG	900
AGGACGGCTC	GCACCACTTC	GTCCGAGCGA	CTCTGACCT	CGATCCCGGG	TGATTCAGCG	960
TGACCGCAGT	GAAATGGAAC	GAACCCCTCA	ACCTGACTA	CGTGGAACTT	GCCTCGAGCG	1020
GTGNTGCA	ACGCGCTTCT	GAAGGTCTC	AAAGGNA			1058

(2) INFORMATION FOR SEQ ID NO:15:

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAATTCGGCA CGAGAGGTGA TCGACATCAT CGGGACCAGC CCCACATCCT GGGAAACAGGC	60
GGCGGGCGGAG GCGGTCCAGC GGGCGCGGGA TAGCGTCGAT GACATCCGCG TCGCTCGGGT	120
CATTGAGCAG GACATGGCCG TGGACAGCCG CGGCAAGATC ACCTACCGCA TCAAGCTCGA	180
AGTGTCGTTT AAGATGAGGC CGGCGCAACC GCGCTAGCAC GGGCCGGCGA GCAAGACGCA	240
AAATCGCAGC GTTTGCGGTT GATTGCTGCG ATTTTGTGTC TGCTCGCCGA GGCCTACCAG	300
GCGCGGCCCA GGTCCGCGTG CTGCCGTATC CAGGCGTGCA TCGCGATTCC GCGCGCCACG	360
CCGGAGTTAA TGCTTCGCGT CGACCCGAAC TGGGCGATCC GCGGNGAGC TGATCGATGA	420
CCGTGGCCAG CCCGTGCGATG CCGAGTTGC CCGAGGAAAC GTGCTGCCAG GCCGGTAGGA	480
AGCGTCCGTA GCGGCGCGTG CTGACCGGCT CTGCCTGCGC CCTCAGTGCG GCCAGCGAGC	540
GG	542

(x) INFORMATION FOR SEQ ID NO:16:

1. SEQUENCE CHARACTERISTICS:

(A) LENGTH: 913 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CGGTCCCGCC CGCGCCTCCG TTGCCCCCAT TGCGGCGGTC GCGGATCAGC TGCGCATCGC	60
CACCATCACC GCGTTTGGCG CGGGCACCAG CGGTGGCGCC GGGGCGCGCG ATGCCACCGC	120
TTGACCGTGG CGCGCGGCGG CGCCATTGCC ATACAGCACC CGCGCGGGGG CACCGTTACC	180
GCGGTCCGCA CGGTGCGCGG CGGTGCGGTT TCAGGCGGGG GAGGCGGAAT GAACCGCGCG	240
CAAGCGCGCC CGCGGACCGG TTGCGCGCTT TTCCGCGCGC CGCGCGCGCG CCGCCAATTG	300
CCGAACAGCC AMGCACCGTT CGCGCCAGCC CGCGCGCGGT TAACGCGCGT GCGGGGCGCG	360
CGCGCGCGAC CGCGCATTAC CGCGCTTCCC GTTCGGTGCC CGCGCGTTAC CGCGCGCGCG	420
GTGTTGCGCG AATATTGGG CGCGACCGCG AGACCGCGCG GGGCCACCAT TGCGCGCGGG	480
CAAGCGAAAC ACAGCGCAAC GGTGCGCGCG GCGCGCGCGT TTGCGCGCAT CACCGGCGAT	540
TCACCGCGAG CACCGCGGTT AATGTTTATG AACCGGTTAC CGCGAGCGCG GCGCGTATTG	600
CGCGCGCGCG GAGNGCGTGG CGCGCGCGCG CGCGAACGCG CAAAAGCGCG GGGTTGCCAC	660
CGCGCGCGCG GGACCGACCG GTCGCGCGCG TCGCGCGGTT CGCGCGGTTG CGCGCGCGCG	720
TTGTTGCGGT CAAGCGGTTA CGCGCGGTTG CGCGGTTTGG CGCGGTTGCG CGCGCGCGCG	780
TTGTTGCGGT GTGCGGTA AGCGAGCGCG CGGTGGGCGT GTGCGCGCGA TTGCGCGCAT	840
TTGTTGCGGT CGCGCGATTG GTGCGGTTGG CGCGCGCGCG GTGCGGTTGG CGCGCGCGCG	900
TTGTTGCGGT CGC	913

(x) INFORMATION FOR SEQ ID NO:17:

1. SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1972 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

GTGATGGTTG	CTGAGCGTGC	TGGCTGCCGT	CGGGCTGGGC	CTGGCCACGG	CGCCGGCCCCA	180
GGCGGCCCCG	CGGGCTTGT	CGCAGGACCG	GTTCGCGGAC	TTCCCCCGGC	TGCCCCCTCGA	240
CCCGTCCGCG	ATGGTCCGCC	AAGTGGCGCC	ACAGGTGGTC	AACATCAACA	CCAAACTGGG	300
CTACAACAAC	GCCGTGGGCG	CCGGGACCGG	CATCGTCATC	GATCCCAACG	GTGTCTGTCT	360
GACCAACAAC	CACGTGATCG	CGGGCGCCAC	CGACATCAAT	GCGTTCAGCG	TGCGTCCCGG	420
CCAAACCTAC	GCGGTGCGATG	TGGTCGGGTA	TGACCGCACC	CAGGATGTCT	CGGTGCTGCA	480
GCTGCGCGGT	GCCGCTGGCC	TGCCGTCCGC	GGCGATCGGT	GGCGGCGTCT	CGGTGCTGGA	540
GCCCGTCTGC	GCGATGGGCA	ACAGCGGTGG	GCAGGGCGGA	ACGCCCCGTG	CGGTGCTTGG	600
CAGGGTGGTC	GCGCTCGGCC	AAACCGTGCA	GCGCTCGGAT	TCGCTGACCG	GTGCCGAAGA	660
GACATTGAAC	GGGTTGATCC	AGTTCGATGC	CGCAATCCAG	CCCGGTGATT	CGGGCGGGCC	720
CGTCGTCAAC	GGCCTAGGAC	AGGTGGTCCG	TATGAACACG	GCCGCGTCCG	ATAACTTCCA	780
GCTGTCCGAG	GGTGGGCAGG	GATTGCGCAT	TCCGATCGGG	CAGGCGATGG	CGATCGCGGG	840
CCAAATCCGA	TGCGGTGGGG	GGTCACCCAC	CGTTCATATC	GGGCTTACCG	CCTTCCTCGG	900
CTTGGGTGTT	GTGACAACA	ACGGCAACGG	CGCACGAGTC	CAACGCGTGG	TCCGAAGCGC	960
TCCGGCGGCA	AGTCTCGGCA	TCTCCACCGG	CGACGTGATC	ACCGCGGTCT	ACGGCGGTCTC	1020
GATCAACTCG	GCCACCCCGA	TGGCGGACGC	GCTTAACGGG	CATCATCCCG	GTGACGTCAT	1080
CTCGGTGAAC	TGGCAAACCA	AGTCGGGCGG	CACGCTTACA	GGGAACGTGA	CATTGGCCGA	1140
GGGACCCCGG	TCCTGATTTC	TCCCGGATAC	CACCGGCGCG	CCGGCCAAAT	GGATTGGCGC	1200
CAGCGCTGAT	TGCCCGCTGA	CCCGCCGAGT	TCCGTCTCCC	GTGCGCGTGG	CATTGTGGAA	1260
GCAATGAACG	AGGCAGAACA	CAGCGTTGAG	CACCGTCCCG	TGCAGGGCAG	TTACGTCCGA	1320
GGCGGTGTGG	TGAGCATCC	GGATGCCAAG	GACTTCGGCA	GCGCGCGCGC	CCTGCCCGCC	1380
GATCCGACCT	GGTTTAAGCA	CGCGGTCTTC	TACGAGGTGC	TGGTCCGGGC	GTTCCTTCGAC	1440
GCCAGCGCGG	ACGGTTCCGN	CGATCTGCGT	GGACTCATCG	ATCGCCCTCGA	CTACCTGCAG	1500
TGGCTTGCCA	TGCACTGCAT	CTGTTGCCGC	CGTTCCTACG	ACTCACCGCT	GCGCGACGGC	1560
GGTACGACA	TTCCGCACTT	CTACAAGGTG	CTGCCCGAAT	TCGGCACCGT	CGACGATTTT	1620
GTGCGCTGG	TGACACGCG	TCACCGGCGA	GGTATCCGCA	TCATCACCGA	CCTGGTGATG	1680
AATCACACCT	CGGASTCGCA	CGCGTGGTTT	CAGGAGTCCC	GCCGCGACCC	AGACGGACCG	1740
TACGGTGACT	ATTACGTGTG	GAGCGACACG	AGCGAGCGGT	ACACCGACCG	CCGGATCATC	1800
TTGCTCGACA	CCGAAGAGTG	GAAGTGGTCA	TTGATCTCTG	TCCGCGGACA	GTNCTACTG	1860
GCAACGATTC	TT					1872

2. INFORMATION FOR SEQ ID NO:13.

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 1461 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:13

TTTCTGCGAA	ACGTGATGCG	GAGGAACAG	TTTCTGCGGT	GAGGCGGAGG	GCCTCCGAGC	
CGCTGCTGCT	CGCCGAGATG	AAGCAATGCG	TTGATGCGAG	AAAAGGCTTG	ACCAACCTGC	120
AAGTACGCTG	CGCAACAAAC	GGAAGAGTGC	ACAGCTTCTT	TGCTATTACG	AGTCCCGATG	180
TGACGCTGCG	CGCCAAATCG	CTGCGGCAAA	ACCGGCTATG	CACCTACAAAC	GACGAGCAGG	240
TTTCTGCTGT	TGCGGTACAA	GGGACAAAC	TGCGGTGAA	ACTGTTGGAU	GACTGGAGCA	300
ATCTGCTGTC	GATTCTGTAA	TTCTTAATTT	CACCGCTGCT	CGATCTGCGC	GCTGGGGTGA	360
GGCAGCTGCT	GTCCGCTGTC	ACGAACCTGC	AAGCGCAAGG	TACGGAAGTG	ATAGACCGAA	420
TTTCAACCA	CAAAATCACC	GGGACATATC	CGCGGAGTGC	TGTCAAGATG	TTTGATCTCT	480
GCCTCAAGAG	TGCAAGCTCT	GGAAGCTCTT	CGATCTGCGA	GGAGGCTCTG	TACCACTCTG	540
TCCTCAAGAG	CATCAAGCTT	CGATCTGCGA	CGATCTGCGT	TACCACTCTG	TACCACTCTG	
AACTCAAGAG	CTCAAGCTCT	CGATCTGCGA	CGATCTGCGT	TACCACTCTG	TACCACTCTG	

CGGTCGATCA	TGGCGGCGAGC	AACGACGTGG	TCGCGCGCCGA	AAACCTCGCC	CCACCGGCGG	900
AAGGCCATTAT	TGGACGTGAC	GATCAAGCTG	GCCCGGTCAT	ACCGGGAGGA	CACCAGCTGG	960
AAGAAGAGGT	TGGCGGCGCTC	GGGCTCAAAC	GGAATGTAAC	CGACTTCGTC	AACCACCAGG	1020
AGCGGATAGC	GGCCAAACCG	GGTGAGTTCC	GCGTAGATGC	GCCCCGGCGT	GTGAGCGCTC	1080
GCGAACCGTG	CTACCCATTG	GGCGGCGGTG	GCGAACAGCA	CCCGATGACC	GGCCTGACAC	1140
GCGCGTATCG	CCAGGCGGAC	CGCAAGATGA	GTCTTCCC	TGCCAGGCGG	GGCCCCAAAA	1200
CACGACGTTA	TCGCGGGCGG	TGATGAAATC	CAGGGTGCCC	AGATGTGCGA	TGGTGTGCGG	1260
TTTGAGGCCA	CGAGCATGCT	CAAAGTCGAA	CTCTTCCAAC	GACTTCCGAA	CCGGGAAGCG	1320
GGCGGCGCGG	ATGCGGCCCT	CACCACCATG	GGAATCCCGG	GCTGACACTT	CCCGCTGCAG	1380
GCAGGCGGCC	AGGTATTCTT	CGTGGCTCCA	GTTCTCGGCG	CGGGCGCGAT	CGGCCAGCGG	1440
GGACACTGAC	TCACGCAGGG	TGGGAGCTTT	CAATGCTCTT	GT		1482

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 876 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GAATTCGGCA	CGAGCGCGCG	ATAGCTTCTG	GGCCCGCGCC	GACCAGATGG	CTCGAGGGTT	60
CGTGCTGGG	GCCACCGCGG	GGCGCACCAC	CCTGACCGGT	GAGGGCGTGC	AACACGCCGA	120
CGGTCACTCG	TTGCTGCTGG	ACGCCACCAA	CCCGGCGGTG	GTTGCCTACG	ACCGGCGCTT	180
CGCGTACGAA	ATCGGCTACA	TCGNGGAAAG	CGGACTGGCC	AGGATGTGCG	GGGAGAACCC	240
GSAGAACATC	TTCTTCTACA	TCACCGTCTA	CAACGAGCCG	TACGTGCAGC	CGCGGAGGCC	300
GSAGAACTTC	GATCCCGAGG	GCCTGCTGGG	GGGTATCTAC	CGNTATCAGG	CGGCCACCGA	360
GSAACGCACC	AACAAGGNGC	AGATCCTGGG	CTCGGGGGTA	GCGATGCCCC	CGGCGCTGCG	420
GGCAGCACAG	ATGCTGCGCG	CCGAGTGGGA	TGTGCGCGCC	GACGTGTGGT	CGGTGACCAG	480
TTGGGGCGAG	CTAAACCGCG	ACGGGTGGGT	CATCGAGACC	GAGAAAGCTC	GCCACCGCGA	540
TCGGCGCGCG	GGCGTGCCCT	ACGTGACGAG	AGCGCTGGAG	AATGCTCGGG	GCCCGGTGAT	600
TCGGGTGTCT	GACTGGATGC	GGCGGCTGCG	CGAGCAGATG	CGACCGTGGG	TGCGGGGCGC	660
ATACCTCAGC	TTGGGCGAGG	ACGGGTTCGG	TTTTTCGGAG	ACTCGGCGCG	CGGTCGCTCG	720
TTACTTCCAC	ACCGACGCGT	AATCGCAGGT	TGGTCCGCGT	TTTGGGAGGG	GTTGGCGCGG	780
TCGACGGGTC	AATATCGACT	CATTGCGTGC	CGGTCTGGGG	TCGCGCGCGC	AGTTACCGCG	840
ATTTCAGCAA	GCTGGGGGGT	TGCGCGCGAG	TAGTT			876

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1071 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATCGGCGCGG	GCTGACGAGAA	TTGCGGAGCA	GAGACAAAAT	TCTAAGCTTT	AATGACGAGAA	60
TAGATTATA	AGCAATTGAC	AGCGGAGAAA	GAATATGTCT	AGATGCGG	TTATTTGAG	120
ATCGGAGAA	TCGCGGAGCT	TCGAGGAAAT	TTTTCAGAG	TCGCGGAG		
TATGCAATG	TCGCGGAGCT					

TTGATGGCAA	CCCTGGTGGC	GGTTGCCGAT	CGGGCCGGGG	CCAACCTGTT	CGAGCTAGAG	480
AACCTTCGTC	CACGTGAAGT	GGATGTGGCG	CCGGCCGCAT	CAGGCGCCCC	GCACGCTGCC	540
GGGGGCCGCC	TCTAGATCCC	TGGGGGGGAT	CAGCGAGTGG	TCCCSTTCGC	CCGCCCGTCT	600
TCCAGCCAGG	CCTTGGTGGC	CCCGGGGTGG	TGAGTACCAA	TCCAGGCCAC	CCCGACCTCC	660
CGGNAAAAGT	CGATGTCCCT	GTAATCATCG	ACGTTCCAGG	AGTACACCGC	CCGGCCCTGA	720
GCTGCCGAGC	GGTCAACGAG	TTGCCGATAT	TCCTTTAACC	CAGGCAGTGA	GGGTCCCAAG	780
GCGGTTGGCC	CGACCGCCGT	GGCCGCACTG	CTGGTCAGGT	ATCGGGGGGT	CTTGCGCGAGC	840
AACAACGTCG	GCAGGAGGGG	TGGAGCCCGC	CGGATCCGCA	GACCGGGGGG	GCGAAAACGA	900
CATCAACACC	GCACGGGATC	GATCTGCGGA	GGGGGGTGCG	GGAAATACCGA	ACCGGTGTAG	960
GAGCGCCAGC	AGTTGTTTTT	CCACCAGCGA	AGCGTTTTTC	GGTCATCGGN	GGCNNTTAAG	1020
T						1021

(2) INFORMATION FOR SEQ ID NO:21:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 321 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CGTCCCGAGC	AACGGAAGAA	CACAACCATG	AAGATGGTGA	AATCGATCGC	CGCAGGTCTG	60
ACCGGCGCGG	CTGCAATCGG	CGCCGCTGCG	GCCGGTGTGA	CTTCGATCAT	GGCTGGCGGN	120
CGCGTCTGAT	ACCAGATGCA	GCCGGTCTCT	TTCCGCGCGC	CATCGCCGTT	GGACCCGGNA	180
TCCGCCCCCTG	ANGTCCCGAC	CGCCGCCGAG	TGGACCAGNC	TGCTCAACAG	NCTCGNCGAT	240
CCCAACGTGT	CGTTTGNGAA	CAAGGGNAGT	CTGGTCCAGG	GNGGNATCGG	NNGNANCGAG	300
GNGNGNATC	GNCANCAACA	A				321

(2) INFORMATION FOR SEQ ID NO:22:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 373 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TTTATCCGCT	TCCGGTTGGG	GACCGGTATT	GGNGCGGGGT	CTTTAACCCG	CTCGGCTAGG	60
CTATCCACCG	GTCGGGAGAG	GTCCACTCCG	ATACTCGGGG	CTCGCTGGAG	CTCCAGGCCG	120
TTTGGTCTCT	GNACCGGCAA	GGGTGAAGG	AGCGGTTGNA	GACCGCGGATC	AAGGCGATTG	180
AGCGGATGAC	CCCGATCCGG	CGCGGCCAGT	GGCAGCTGAT	CTTCGGGGAG	CGCAAGACCG	240
TTTAAACCCG	CGGTCTCTCT	CGACACCAT	CCTCAAACCA	CGCGGAAGAA	CTGGGAGTCT	300
CTTGGATCCC	AAGAAGTAGG	TCCGCTTCTG	TATACCTTGG	GCATCGGGCA	AGAAGGGGAA	360
TTTAACTCCG	CCG					373

(2) INFORMATION FOR SEQ ID NO:23:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 321 base pairs
- (B) TYPE:

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GTGACGCCGT GATGGGATTC CTGGGCGGGG CCGGTCCGCT GGUGGTGGTG GATCAGCAAC	60
TGGTTACCCG GGTGCCGCAA GGCTGGTCGT TTGCTCAGGC AGCCGCTGTG CCGGTGGTGT	120
TCCTTGACGC CTGGTACGGG TTGGCCGATT TAGCCGAGAT CAAGGCGGGC GAATCGGTGC	180
TGATCCATGC CGGTACCGGC GGTGTGGGCA TGGCGGCTGT GCAGCTGGCT CGCCAGTGGG	240
GCSTGGAGGT TTTCTGCACC GCCAGCCGTG GNAAGTGGGA CACGCTGCGC GCCATNGNGT	300
TTGACGACGA NCCATATCGG NGATTCCNC ACATNCGAAG TTCCGANGGA GA	352

(2) INFORMATION FOR SEQ ID NO:24:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 726 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GAAATCCGGG TTCATTCCGT TCGACCAGCG SCTGGCGATA ATCGACGAAG TGATCAAGCC	60
GCGGTTCGCG GCGTCATGG GTCACAGCGA GTAATCAGCA AGTCTCTG TATATCGCAC	120
CTAGGTCCCA GTTGCTTGCC AGATCGCTTT CGTACCGTCA TCGCATGTAC CCGTTCGCGT	180
GCCGACGCT CATGCTGGCG GCGTGCATCC TGGCCACGGG TGTGGCGGGT CTCGGGGTCC	240
GCGGSCAGTC CGCAGCCCAA ACCGCGCCCG TGGCCGACTA CTACTGGTGC CCGGGGCAGC	300
CTTTGACCC CGCATGGGGG CCCAACTGGG ATCCCTACAC CTGCCATGAC GACTTCCACC	360
GCGACAGCGA CGGCCCCGAC CACAGCCGCG ACTACCCCGG ACCCATCCTC GAAGGTCCCG	420
TGCTTGACGA TCCCGGTGCT GCGCCGCGCG TCCCGGCTGC CGGTGGCGGG GCATAGCCT	480
CGTTGACCGG GCGGCTCAG CGAATACCG TATAAACCGG GCGGTGCCCG CCGCAAGCTA	540
CGACCCCGCG CGGGGCAGAT TTACGCTCCG GTGCGGATGG ATCGCCCGCG CCGATGACAG	600
AAAATAGGG ACSGTTTTGG CAACCGCTT GAGGACGCTT GAAGGGAACC TGTCATGAAC	660
CGGACAGCG CCTCCACCAT CGACATCGAC AAGGTTGTTA CCGGCACACC CGTTCGCGCG	720

(2) INFORMATION FOR SEQ ID NO:25:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 680 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:25:

AGGATGAGG AGGAACGTGG GGGGACGAG CCGTATGCG TATATGCAAG CGACCGGCAT	60
GCTGCGCGAG CATATCCAAAG CATGCTGGGT GCGCACTGAG CGACCTTTTG ACCAGCCCGG	120
CTGCGGATG GCGGCGCGGT GAAGTCAAT GCGGCGCGCT TGTGCACTG ATGAACCGGA	180
ATAGGGAAGA ATAGCGCGGT GATTTGGGAG TTGAATGTGG GTATGCGTG GAAATCCAAAT	240
GCGGCGGCAT GCTCGCGCGC GACGAGGTC GCGGAGGCGG GCGAGGCTGA ATCTGGAGGG	300
AGCACTCAAT GCGGCGCATG AAGGCTCGGA CCGGCGACCG TCGTTTGGAA GCAACTAAGG	360
AGGCGCGCGG CATTTGCGAT CGAGTATTA TTGAAGGTGG CGGTGCTTG CTCGTGAGG	420
TGACACCGGA CGAAGCGCGG CGACTGCTG AGGAAGTCAA AGTCTTATG AGTCTATG	480

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 160 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

```

AACGGAGGCG CCGGGGGTTT TGGCGGGGCC GGGGCGGTCG GCGGCAACGG CCGGGCCGGC      60
GGTACCGCCG GGTGTGTCGG TGTCCGCGGG GCCGGTGGGG CCGGAGGCAA CCGCATCGCC      120
GGTGTACCGG GTACGTCGGC CAGCACACCG GGTGGATCCG                               160
  
```

(2) INFORMATION FOR SEQ ID NO:27:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 172 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

```

GACACCGATA CGATGGTGAT GTACGCCAAC GTTGTGACA CGCTCGAGGC GTTCACGATE      60
CAGCCGCACAC CCGACGGCGT GACCATCGGC GATCGGGCCC CGTTCGCGGA GGCGGCTGCC      120
AAGGCGATCG GAATCGACAA GCTGCGGGTA ATTCATACCG GAATGGACCC CGTCGTCGCT      180
GAACGCGAAC AGTGGGACGA CCGCAACAAC ACGTTGGCGT TGGCGCCCGG TGTCGTTGTC      240
GCCTACGAGC GCAACGTACA GACCAACGCT CG                               272
  
```

(2) INFORMATION FOR SEQ ID NO:28:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 317 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

```

GCAGCGGGTG GTTCTGGGAC TATGTGCGCA CGGTGAGGCA GCGGACGTC CCGGAGGTGA      60
AGCGGATGGA GCAGAGGAT TCGGTGCGCG GTTGTATGCG GTACGTGGCC GGTATCACCG      120
GCGAGGAGGT GAAGTGGGT GAACCGGCGC GGGTCATCGG GTGCGAGCGG GCGACGATCC      180
GTTCGGATCT GCGGTGTTG GAGACGCTCT ATCTGTACA TCGGTGCGC GCGTGGTCGC      240
GGAATCTGAC GCGGAAGATC AGAAGCGCT DAAAGATGTA GTGTGTGAC AGTGGCTTCG      300
CGGCTTGCTT GCGCGGGG                               317
  
```

(2) INFORMATION FOR SEQ ID NO:29:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 181 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single

GATCGTGGAG CTGTGGATGA ACAGCGTTGC CGGACGCGCG GCGGCCAGCA CSTCGGTGTA 60
 GCAGCGCCGG ACCACCTCGC CGGTGGGCAG CATGGTGATG ACCACGTCCG CCTCGGCCAC 120
 CGCTTCGGGC GCGCTACGAA ACACCGCGAC ACCGTGCGCG GCGGCGCCGG ACGCCGCCGT 180
 GG 182

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 308 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GATCGGGAAG TTTGGTGAAC AGGTGGTGA CCGGAAAGTC TGGGCGCCTC CGAAGCGGGT 60
 CGGCGTTTAC GAGGCGAAGA CACGCTGTGC CGAGCTGCTG CGGCTGCTCT ACGGCGGGCA 120
 GAGGTTGAGA TTGCGCGCGG CGGCGAGCGG GTAGCAAAGC TTGTGCGGCT GCATCCTCAT 180
 GAGACTCGGC GGTTAGGCAT TGACCATGGC GTGTACCGCG TGCCCGACGA TTIGGACGCT 240
 CGTTGTGAG ACCACGTGCT CGAAGGCTTT CACCGGTGAA GCGCTACCTC ATCGACACCC 300
 AGGTTTGG 308

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 267 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CGGACGACGA GCAACTCAGG TGGATGATG TCGGCGAGCG CATGAGGAGC GGAGAGAATC 60
 CGGCGGAAGC TCGGCGCGG CAGTCTCTCA TATGACCGG CGGTAGAGGG CTCGCCCGAT 120
 CGGACGCGAC TATTCTGGTG TCGGCTTGGC CGGTAAGAGC GGGTAAAAGA ATGTGAGGGG 180
 ACAGSATGAG CAATCAGAGC TACCGAGTGA TCGAGATCGT CGGACCTCG CCGACGGCG 240
 TCGACGCGGC AATCCAGGGC GGTCTGC 267

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1544 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CTGCTCCGA AAGAATGTGA GGGGACGGA TGACCAATCA CAGCTACCGA GTGATCGAGA 60
 TTTTGGGAC CTGCTCCGA TCGCTGAGG GCGCAATCCA GGGGCTCTG GCGGAGGCTG 120
 TTTTGGGAC CTGCTCCGA TCGCTGAGG GCGCAATCCA GGGGCTCTG GCGGAGGCTG 180
 TTTTGGGAC CTGCTCCGA TCGCTGAGG GCGCAATCCA GGGGCTCTG GCGGAGGCTG 240

GCTCGACCTG	TGTGGGCTGC	AGCCGGACGA	AGCGGTGCTC	GAUGTCGGCT	GCGGCTCGGG	480
GCGGATGGCG	TTGCCGCTCA	CCGGCTATCT	GAACAGCGAG	GGACGCTACG	CCGGCTTCGA	540
TATCTCGCAG	AAAGCCATCG	CGTGGTGCCA	GGAGCACATC	ACCTCGGCGC	ACCCCAACTT	600
CCAGTTCGAG	GTCTCCGACA	TCTACAACTC	GCTGTACAAC	CCGAAAGGGA	AATACCAGTC	660
ACTAGACTTT	CGCTTTCCAT	ATCCGGATGC	GTCGTTGAT	GTGGTGTTC	TTACCTCGGT	720
GTTCAACCCAC	ATGTTTCCGC	CGGACGTGGA	GCACTATCTG	GACGAGATCT	CCCGCGTGCT	780
GAAGCCCGGC	GGACGATGCC	TGTGCACGTA	CTTCTTGCTC	AATGACGAGT	CGTTAGCCCA	840
CATCGCGGAA	GGAAAGAGTG	CGCACAATT	CCAGCATGAG	GGACCGGGTT	ATCGGACAAT	900
CCACAAGAAG	CGGCCCGAAG	AAGCAATCGG	CTTGCCGGAG	ACCTTCGTCA	GGGATGTCTA	960
TGGCAACTTC	GGCCTCGCCG	TGCACGAACC	ATTGCACTAC	GGCTCATGGA	GTGGCCGGGA	1020
ACCACGCCTA	AGCTTCCAGG	ACATCGTCAT	CGCGACCAAA	ACCGCGAGCT	AGGTCGGCAT	1080
CCGGGAAGCA	TCGCGACACC	GTGGCGCCGA	GCGCCGCTGC	CGGCAGGCCG	ATTAGGCGGG	1140
CAGATTAGCC	CGCCCGGGCT	CCCGGCTCCG	AGTACGGCGC	CCCGAATGGC	GTCACCGGCT	1200
GGTAACCACG	CTTGCGCGCC	TGGGCGCGCG	CCTGCCGGAT	CAGGTGGTAG	ATGCCGACAA	1260
AGCCTGCGTG	ATCGGTCACT	ACCAACGGTG	ACAGCAGCCG	GTTGTGCACC	AGCGCGAACG	1320
CCACCCCGGT	CTCCGGGTCT	GTCCAGCCGA	TCGAGCCGCC	CAAGCCCAAC	TGACCAAAAC	1380
CCGGCATCAC	GTGCGCGATC	GGCATACCGT	GATAGCCAAG	ATGAAAATTT	AAGGGCACCA	1440
ATAGATTTCC	ATCCGECAGA	ACTTCCGCTC	GGTTCCGGGT	CAGGCCCGTG	ACCAGCTCCC	1500
CCGACAGAA	CCGTATGCCG	TCGATCTCCG	CTCGTGCCG			1539

(12) INFORMATION FOR SEQ ID NO:33:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 351 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(13) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CTCCAGGGTG	GCSTGGATGA	GCSTCACCCG	GGGGCAGGCC	GAGCTGACCG	CCGCCCCAGGT	60
CCGGGTTGCT	GCGGCGGCCCT	ACGAGACGSC	GTATGGGCTG	ACGGTGCCCC	CCCGCGGTGAT	120
CCCGGAGAAC	CGTGCTGAAC	TGATGATTCT	GATAGCGACC	AACCTGTTGG	GGCAAAACAC	180
CCCGCGGATC	GCGGTCAACG	AGGCCGAATA	CCCGGAGATG	TGGGCCCAAG	ACGCCCCCCC	240
GATGTTTGGC	TACGCGCCCG	CGACCGCGAC	GGCGACGGCG	ACGTTGCTGC	CGTTCGAGGA	300
GCGCGCGGAG	ATGACCAGCG	CGGTTGGGCT	CCTCGAGCAG	CCCGCCCGCG	TGGAGGAGGC	360
CTCCGACACC	GCCCGCGCGA	ACCAAGTTGAT	GAACAATGTC	CCCGAGGCCG	TGAAACAGTT	420
GGCCGAGCCC	ACGCGAGGCA	CCAGGCTTCG	TTCCAAGCTG	TGTGGCCTGT	GGAAGACGGT	480
CTCCCGGCAT	CGGTGCGCGA	TCAGCAACAT	GCTGTGATG	CCCAACAACC	ACATGTGAT	540
GACCAACTCG	GCTGTGTCGA	TGACCAACAC	CTTGAGCTCG	ATGTTGAAGG	GCTTTGCTCG	600
CCCGGCGGCC	GCCGAGGCCG	TGCAAACTCG	CCCGCAAAAC	GGGGTCCGGG	CGATGAGCTG	660
CTTGGGCGAG	TGGCTGGGTT	CTTGGGCTCT	GGGGGTGCGG	TTGGGCGCGA	ACTTGGGTCC	720
CGGGGCTCC	GTACGCTATG	CTTAAAGGGA	TGGGGGAAAA	TATGCAAAAT	CTGGTGGGCG	780
GAAGGCTGCT	CCCGGCTAAG	CTTAAAGGGA	CTTTTCTGGA	TGGGTGAAC	TTGTTCAACG	840
GAAGAGTTA	C					851

(14) INFORMATION FOR SEQ ID NO:34:

(1) SEQUENCE CHARACTERISTICS:

- A LENGTH: 351 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single

GATCGATCGG	GCGGAAATTT	GGACCAGATT	CGCCTCCGGC	GATAACCCAA	TCAATCGAAC	60
CTAGATTTAT	TCCGTCCAGG	GGCCCCAGTA	ATGGCTCGCA	GGAGAGGAAC	CTTACTGCTG	120
CGGGCACCTG	TCGTAGGTCC	TCGATACGGC	GGAAGGCGTC	GACATTTTCC	ACCGACACCC	180
CCATCCAAAC	GTTCGAGGGC	CACTCCAGCT	TGTGAGCGAG	GCGACCGAGT	CGCAGGCTGC	240
GCTTGGTCAA	GATC					254

(2) INFORMATION FOR SEQ ID NO:35:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1227 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GATCTCTGAC	GAAGCGGCG	CCGCCAAGCG	GAAGTCGCTG	TTGGACCGAG	AGGGACCGGA	40
CGATCTCGCG	GTGCGGATCG	CGTTTCAGCG	GGGGGGCTGC	GCTGGATTGC	GCTATAACCT	120
TTTCTTCGAC	GACCGGACCG	TGGATCGTGA	CCAAACCGCG	GAGTTCGGAT	GTGTCAAGTT	180
GATCGTGGAG	CGGATGACCG	CGCCGTATGT	GGAAGGCGCG	TCGATCGGATT	TCGTCGACAC	240
TATTGAGAAG	CAAGGTTTAC	CATCGACAAT	CCCAACGCCA	CCGGCTCCTG	CCGGTGCGGG	300
GATTCGTTCA	ACTGATAAAA	CGCTAGTACG	ACCCCGCGGT	GCGCAACAGG	TACGAGCACA	360
CCAAGACCTG	ACCGCGCTGG	AAAAGCAACT	GAGCGATGCC	TTGCACCTGA	CCGCGTGGCG	420
GGCGCGCGCG	GCGAGGTGTC	ACCTGCATGG	TGAACAGCAC	CTGGGCCTGA	TATTGCGACC	480
AGTACACGAT	TTTGTGGATC	GAGGTCACTT	CGACCTGGGA	GAAGTCTCTG	CGGAACCGCT	540
CGCTGCTCAG	CTTGGCCAAAG	GCTGTATCGG	AGCGCTTTGC	GCGCACGCGG	TCGTGGGATAC	600
CGCACAGCGC	ATTGCGAAGC	ATGGTGTCCA	CATCGCGGTT	CTCCAGCGCG	TTGAGGTATC	660
CCTGAATCGC	GGTTTTGGCG	GGTCCCTCCG	AGAATGTGCC	TGCCGTGTTG	GCTCCGTTGG	720
TCCGACACCC	GTATATGATC	GCCGCCCTCA	TAGCCGACAC	CAGCGCGAGG	GCTACCACAA	780
TGCCGATCAG	CAGCGGCTTG	TGCCGTCCGT	TCCGGTAGGA	CACCTGCGGC	GGCACGCGCG	840
GATATGCGGC	GGGCGCGAGC	CCCGCGTCCG	CTGCCGGTTC	CGGGGCGAAG	GCCGGTTCCG	900
GGGCGCGGAG	GTGCTGGGGG	TAGTCCAGCG	CTTGGGGTGC	GTGGGATGAG	GGCTCTGGGT	960
ACGCGCGCGG	TCCGTTGGTG	CCGACACCGG	GCTTCCGCGA	TTGGGACCG	GCCATTGTGG	1020
TTCTGCTAGG	GTGGTGGAGC	GGACCAAGCTG	CTAGGGCGAC	AACCGCCCGT	CGCGTCAGCC	1080
GGCAGCATCG	GCAATTCAGGT	GAGCTCCCTA	GGCAGGCTAG	TGCACACAGT	CGCGTCAGCT	1140
CTCAACCGCA	CGGGCGCGGG	CGCGCGCGCG	ATAATGTTGA	AAGAATAGGG	AACCTTAGGA	1200
ACGAAGGACG	GAGATTTTGT	GACGATC				1260

2 INFORMATION FOR SEC IN NORMAL

SEQUENCE CHARACTERISTICS

- A LENGTH: 181 base pairs
B TYPE: nucleic acid
C STRANDEDNESS: single
D TOPOLOGY: linear

X1. SEQUENCE DESCRIPTION: SEQ ID NO: 36.

1. THE UNITED STATES OF AMERICA
 2. DEPARTMENT OF JUSTICE
 3. DIVISION OF INVESTIGATION
 4. WASHINGTON, D. C. 20535
 5. MAY 19 1964
 6. MEMPHIS, TENNESSEE
 7. TO: SAC, MEMPHIS
 8. FROM: SAC, NEW YORK
 9. SUBJECT: MARTIN LUTHER KING, JR.;
 10. CIVIL RIGHTS; RACIAL MATTERS;
 11. RE: NEW YORK TELETYPE TO BUREAU
 12. MAY 18, 1964.

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 290 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GCGGTGTCGG CGGATCCGGC GGGTGGTTGA ACGGCAACGG CGGTGTCGGC GGCCGGGGCG	60
GCGACGGCGT CTTTGCCGGT GCCGGCGGCC AGGGCGGCCT CGGTGGGCAG GGCGGCAATG	120
GCGGCGGGCTC CACCGCGGC AACGGCGGTC TTGGCGGGCG GGGCGGTGGC GGAGGCAACG	180
CCCCGGACGG CGGCTTCGGT GGCAACGGC GTAAGGGTGG CCAGGGCGGN ATTGGCGGCG	240
GCACTCAGAG CGCGACCGC CTCGGNGGTG ACGGCGGTGA CGGCGGTGAC	290

(2) INFORMATION FOR SEQ ID NO:38:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 34 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GATCCAGTGG CATGGNGGCT CTCAGTGGAA GCAT	34
---------------------------------------	----

(2) INFORMATION FOR SEQ ID NO:39:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 155 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GATGCGTGCT CGTCCCCCCC TTGCGGCGCA CGGCACCGGT CGCAGCGTTA CCGAACAAGC	60
TTGGGTGGTC GCCAGCACCC GCGGCAACCC CGACCGCGGA TTGGAAGCAT GGCAGCGTCS	120
GATGCGTACG ATTGCGGCGG GTCGGAAGGG TACCG	155

(2) INFORMATION FOR SEQ ID NO:40:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 57 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 132 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GATCCACCGC GGGTGCAGAC GGTGCCCCCG GCGCCACCCC GACCAGCGGC GGCAACGGCG	60
GCACCGGCGG CAACGGGCGG AACGCCACCG TCGTCGGNGG GGCCGGCGGG GCCGGCGGCA	120
AGGGCGGCAA CG	132

(2) INFORMATION FOR SEQ ID NO:42:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 132 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GATCGGCGGG CGGNACGGNC GGGGACGGCG GCAAGGGCGG NAACGGGGGC GCCGNAGCCA	60
CCNGCCAAGA ATCCTCCGNG TCNCCAATG GCGGAATGG CGGACAGGGC GGCAACGGCG	120
GCANCGGCGG CA	132

(2) INFORMATION FOR SEQ ID NO:43:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 702 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

CGGCACGAGG ATCGGTACCC TGGGGCATCG SCAGCTGCGG ATTGCGCGGG TTTCCCCACC	60
CGAGGAAAGC CGCTACCAGA TGCGGCTGCC GAAGTAGGGC GATCCGTTTC CGATGCCGGC	120
ATGAACGGGC GGCATCAAAT TAGTGCAGGA ACCCTTCAGT TTAGCGACGA TAATGGCTAT	180
AGCACTAAGG AGGATGATCC CATATGAGCG AGTCCAGAGC CTTGACGGTG GATCAGCAAG	240
AGATTTTGA AAGGGGCAAC TAGGTGAGG TCCGATGGC GGACCGACCG ACTGATGTCC	300
TCATCACACC GTGCGAATC AGGNGGNTA AAAACGCCCC TCAACAGNTG GTNTTGTCCG	360
CGGACAACAT GCGGGAATAC TTGGGGTTCB TTGCAAGA GCGGZAGCGT CTGGCGACCT	420
CGCTGCGCAA CGCGGCCAAG GNGTATGCGG AGGTTGATGA GAGGCTGCG ACCGCGCTGG	480
ACAACGACCG CGAAGGAACT GTGAGGCGAG AATCGGCCGG GCGCGTCGGA GGGGACAGTT	540
CGGCGGAACT AATCGATACG CGGAGGGTGG CGAGGGCTGG TGAACCCAAC TTCATGGATC	600
TCAAAGAAGC GGCAAGGAAG CTCGAAACCG GCGACCAAGG CGCATCCCTC GCGCACTGNG	660
GGGATGGGTG GAACACTTNC ACCCTGACCG TGAAGGGCGA CG	702

(2) INFORMATION FOR SEQ ID NO:44:

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GAAGCCGCGAG CGCTGTCGGG CGACGTGGCG GTCAAAGCGG CATCGCTCGG TGGCGGTGGA	60
GGCGCGCGGG TGCCGTCGGC GCCGTGGGA TCCGCGATCG GGGCGCCGA ATCGGTGCGG	120
CCCCTGCGG CTGGTGACAT TGCCGGCTTA GGCCAGGAA GGGCCGGCG CGGCGCGCG	180
CTGGCGCGG GTGGCATGG AATGCCGATG GGTGCCGCG ATCAGGGACA AGGGGGCGCC	240
AAGTCCAAG GTTCTCAGCA GGAAGACGAG GCGCTCTACA CCGAGGATCC TCGTGCCG	298

(2) INFORMATION FOR SEQ ID NO:45:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1058 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

x1 SEQUENCE DESCRIPTION: SEQ ID NO:45:

CGGCGCGAGG ATCGAATCGG GTGCGCGGGA GCACAGCGTC GCACTGCACC AOTGGAGGAG	60
CCATGACCTA CTCGCGGGGT AACCCCGGAT ACCCGCAAGC GCAGCCCGCA GGCTCCTACG	120
GAGCGCTCAC ACCCTCGTTC GCCCACGCGG ATGAGGGTGC GAGCAAGCTA CCGATGTACC	180
TGAACATCGC GGTGGCAGTG CTCGGTCTGG CTGCGTACTT CGCCAGCTTC GGCCCAATGT	240
TCACCGTCAG TACCGAATC GGGGGGGGTG ATGGCGCAGT GTCCGGTGAC ACTGGGCTGC	300
CGGTGCGGGT GGCTCTGCTG GCTGCGCTGC TTGCGGGGT GGTCTGTGGT CCGAAGGCCA	360
AGAGCCATGT GACGGTAGTT GCGGTCTCG GGTACTCGG CATTATTTCTG ATGGTCTCGG	420
CGACGTTTAA CAAGCCGAGC GCCTATTGCA CCGGTTGGGG ATTGTGGGTT GTGTTGCTT	480
TCATCGTGT CCAGCGCGTT GCGGCGAGTC TGCGGCTCTT GGTGGAGACC GGCGCTATCA	540
CGGCGCGGCG GCGCGGCGCC AAGTTGACG CSTATGGACA GTACGGCGCG TACGGGAGT	600
ACGGGCGAGTA CGGCGTGGAG CCGGCTGGGT ACTACGGTCA GAGGGTGTGT CAGCAGGCGG	660
CGGGAGTACA GTGCGCGGCG CGCGAGCAGT CTCGCGAGCG TCCCGGATAT GGTGCGAGT	720
ACGGCGGCTA TTGTTGAGT GCGAGCGAAT CGGGAGTGG ATACAATGCT CAGCGCGCTG	780
CGGCGCGGCG GCGCGAGTGC GGTGCGCAAC AATCGCACCA GGGCGATGC AGCGGAGCTA	840
CGGCGTTTCC GAGGTTGAGC GCACGACGAC CGGTGAGTGC CGGGAAGGGG TCGCAGGCTG	900
GTTCGGCTCC AGTCAACTAT TCAAAACCCA GCGGGGGCGA GCAGTGTGCG TCGCGCGGG	960
GGGCGCGCGT CTAACCGGGC GTTCCCGCGT CCGGTGCGCG GTGTGCGCGA AGAGTGAACA	1020
GGTGTGAGC AAGCGCGGAC GATGCTCGTG CGGAATTC	1058

(2) INFORMATION FOR SEQ ID NO:46:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 107 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

x1 SEQUENCE DESCRIPTION: SEQ ID NO:46:

CGGCGCGAGG ATCGAATCGG GTGCGCGGGA GCACAGCGTC GCACTGCACC AOTGGAGGAG	60
CCATGACCTA CTCGCGGGGT AACCCCGGAT ACCCGCAAGC GCAGCCCGCA GGCTCCTACG	120

(2) INFORMATION FOR SEQ ID NO:47:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 170 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

CGGTCGCGAT GATGGCGTTG TCGAACGTGA CCGATTCTGT ACCGCCGTCG TTGAGATCAA	60
CCAACAACGT GTTGGCSTCG GCAAATGTGC CGNACCCGTG GATCTCGGTG ATCTTGTCT	120
TCTTCATCAG GAAGTGCACA CCGGCCACCC TGGCCTCGGN TACCTTTCGG	170

(2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 127 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GATCGGGCGG CACGGGGGGT GCGGGCGGCA GCACCGCTGG CGCTGGCGGC AACGGCGGGG	50
CGGGGGGTGG CGGCGGAACC GGTGGGTTCG TCTTGGGCAA CGGCGGTGCC GGCGGGCACC	120
GGGCGCGT	127

(2) INFORMATION FOR SEQ ID NO:49:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 91 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

AGCGGCAAG GCGGGACCG GGGGACCG GAGCGGCGG GCGGGCGGCA AGCGGGCA	60
AGCGGCTCG GCGGCAACG	91

(2) INFORMATION FOR SEQ ID NO:50:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 149 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

(2) INFORMATION FOR SEQ ID NO:51:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 355 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CGGCACGAGA TCACACCTAC CGAGTGATCG AGATCGTCGG GACCTCGCCC GACGGTGTCTG	60
ACGCGGNAAT CCAGGGCGGT CTGGCCCGAG CTGCGCAGAC CATGCGCGCG CTGGACTGGT	120
TCGAAGTACA GTCAATTGGA GGCCACCTGG TCGACGGAGC GGTGCGCGAC TTCCAGGTGA	180
CTATGAAAGT CCGCTTCCGC CTGGAGGATT CCTGAACCTT CAAGCGCGGC CGATAACTGA	240
GGTGCATCAT TAAGCGACTT TTCCAGAACA TCCTGACGCG CTCGAAACCC GGTTCAGCCG	300
ACGGTGGGTC CCGCGACCCG CTGCTTCCAA AATCCCTGCG ACAATTCGTC GGCCG	355

(2) INFORMATION FOR SEQ ID NO:52:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 999 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:52:

ATGCATCACC ATCACCATCA CATGCATCAG GTGGACCCCA ACTTGACAGC TCSCAAGGGA	60
CGATTGGCGG CACTGGCTAT CGCGCCGATG GCCAGCGCCA GCCTGGTGAC CGTTGCGGTG	120
CCCGCGACCG CCAACSCCGA TCCGGAGCCA GCGCCCGCGG TACCCACAAC GGCCGCGCTG	180
CCCGCTTCGA CCGCTGCAGC GGCACCCGCA CCGCGGACAG CTGTTCCGCC CCCACGACCG	240
CCCGCCGCA ACACGCGGAA TGCCGAGGCG GCGGATCCCA ACSCAGCACC TCCGCGCGCG	300
ACCCGGAACG CACCGCCGCG ACTTGTGATT GCGCCAAAGC CACCCCAACC TGTCGGGATG	360
GACAACCCGG TTGGAGGATT CAGCTTCGCG CTGCTGTGTC GCTGGGTGGA GTCTGACGCG	420
GCCCACITCG ACTACGGTTC AGCACTCCTC AGCAAAACCA CCGGGGACCC GCCATTTCCC	480
GGACAGCCCG CCGCGGTGGC CAATGACACC CGTATCGTGC TCGGCCGCGT AGACCAAAAG	540
CTTTACGCCA GCGCCGAAGC CACCGACTCC AAGGCCGCGG TCCGCTTGGG CTCGGACATG	600
GGTGAATTCT ATATGCCCTA CCGGGGACCG CGGATCAAG AGGAAACCGT CTCGCTCGAC	660
CCCAACGGGG TGTCTGGAAG CGCTCTGAT TACCAAGTCA AGTTCAAGCA TCCGAGTAAG	720
CCGAACGGGC AGATCTGGAC GGGGTAATC GGTGCGCGG CCGCAAGCG ACCCGACGCG	780
GGGCTGCGTC AGCGCTGGTT TGTGCTATCG CTGGGACCG CCAACAAAGT GGTGACAAG	840
GGGCGGCTCA AGGCGCTGGC CGAATGATC CGGCTTTTGG TCGGCGCGCG CCGCGCGCGC	900
GCACCGGCTC CTGCAGAGCG GGTCTGCGG CCGCGCGCGG CCGGGGAAGT CGCTCCTACC	960
CCGAAGACAC CGACACCCCA CCGGATCTTA CCGGCTTGA	999

(2) INFORMATION FOR SEQ ID NO:53:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 355 base pairs
 - (B) TYPE: nucleic acid

Met His His His His His Met His Gln Val Asp Pro Asn Leu Thr
1 5 10 15
Arg Arg Lys Gly Arg Leu Ala Ala Leu Ala Ile Ala Ala Met Ala Ser
20 25 30
Ala Ser Leu Val Thr Val Ala Val Pro Ala Thr Ala Asn Ala Asp Pro
35 40 45
Glu Pro Ala Pro Pro Val Pro Thr Thr Ala Ala Ser Pro Pro Ser Thr
50 55 60
Ala Ala Ala Pro Pro Ala Pro Ala Thr Pro Val Ala Pro Pro Pro Pro
65 70 75 80
Ala Ala Ala Asn Thr Pro Asn Ala Gln Pro Gly Asp Pro Asn Ala Ala
85 90 95
Pro Pro Pro Ala Asp Pro Asn Ala Pro Pro Pro Pro Val Ile Ala Pro
100 105 110
Asn Ala Pro Gln Pro Val Arg Ile Asp Asn Pro Val Gly Gly Phe Ser
115 120 125
Phe Ala Leu Pro Ala Gly Trp Val Glu Ser Asp Ala Ala His Phe Asp
130 135 140
Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr Gly Asp Pro Pro Phe Pro
145 150 155 160
Gly Gln Pro Pro Pro Val Ala Asn Asp Thr Arg Ile Val Leu Gly Arg
165 170 175
Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu Ala Thr Asp Ser Lys Ala
180 185 190
Ala Ala Arg Leu Gly Ser Asp Met Gly Glu Phe Tyr Met Pro Tyr Pro
195 200 205
Gly Thr Arg Ile Asn Gln Glu Thr Val Ser Leu Asp Ala Asn Gly Val
210 215 220
Ser Gly Ser Ala Ser Tyr Tyr Gln Val Lys Phe Ser Asp Pro Ser Lys
225 230 235 240
Pro Asn Gly Gln Ile Trp Thr Gly Val Ile Gly Ser Pro Ala Ala Asn
245 250 255
Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp Phe Val Val Trp Leu Gly
260 265 270
Thr Ala Asn Asn Pro Val Asp Lys Gly Ala Ala Lys Ala Leu Ala Glu
275 280 285
Ser Ile Arg Pro Leu Val Ala Pro Pro Pro Ala Pro Ala Pro Ala Pro
290 295 300
Ala Ala Pro Ala Pro Ala Ala Ala Ala Gly Glu Val Ala Pro Thr
305 310 315
Pro Thr Thr Pro Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
320 325 330

INFORMATION FOR SEC 17 NO: 54

SEQUENCE CHARACTERISTICS

- A: LENGTH: 20 AMINO ACIDS
E: TYPE: AMINO ACID
I: STRANDEDNESS
T: TOPOLOGY

1 5 10 15
Val Ala Ala Leu
20

(2) INFORMATION FOR SEQ ID NO:55:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Ala Val Glu Ser Gly Met Leu Ala Leu Gly Thr Pro Ala Pro Ser
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:56:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 19 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

Ala Ala Met Lys Pro Arg Thr Gly Asp Gly Pro Leu Glu Ala Ala Lys
 1 5 10 15
 Glu Gly Arg

(2) INFORMATION FOR SEQ ID NO:57:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:58:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 14 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:

Asp Ile Gly Ser Glu Ser Thr Glu Asp Gln Gln Xaa Ala Val
 1 5 10

(2) INFORMATION FOR SEQ ID NO:59:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 13 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

Ala Glu Glu Ser Ile Ser Thr Xaa Glu Xaa Ile Val Pro
 1 5 10

(2) INFORMATION FOR SEQ ID NO:60:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 17 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Asp Pro Glu Pro Ala Pro Pro Val Pro Thr Ala Ala Ala Ala Pro Pro
 1 5 10 15
 Ala

(2) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

Ala Pro Lys Thr Tyr Xaa Ile Ile Lys Lys Gly Thr Asp Thr Gly
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:62:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 187 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

Thr	Gly	Ser	Leu	Asn	Gln	Thr	His	Asn	Arg	Arg	Ala	Asn	Glu	Arg	Lys
1				5										15	
Asn	Thr	Thr	Met	Lys	Met	Val	Lys	Ser	Ile	Ala	Ala	Gly	Leu	Thr	Ala
			20					25					30		
Ala	Ala	Ala	Ile	Gly	Ala	Ala	Ala	Ala	Gly	Val	Thr	Ser	Ile	Met	Ala
		35				40						45			
Gly	Gly	Pro	Val	Val	Tyr	Gln	Met	Gln	Pro	Val	Val	Phe	Gly	Ala	Pro
	50					55					60				
Leu	Pro	Leu	Asp	Pro	Ala	Ser	Ala	Pro	Asp	Val	Pro	Thr	Ala	Ala	Gln
65				70							75				80
Leu	Thr	Ser	Leu	Leu	Asn	Ser	Leu	Ala	Asp	Pro	Asn	Val	Ser	Phe	Ala
			85						90					95	
Asn	Lys	Gly	Ser	Leu	Val	Glu	Gly	Gly	Ile	Gly	Gly	Thr	Glu	Ala	Arg
		100						105					110		
Ile	Ala	Asp	His	Lys	Leu	Lys	Lys	Ala	Ala	Glu	His	Gly	Asp	Leu	Pro
	115					120						125			
Leu	Ser	Phe	Ser	Val	Thr	Asn	Ile	Gln	Pro	Ala	Ala	Ala	Gly	Ser	Ala
	130					135					140				
Thr	Ala	Asp	Val	Ser	Val	Ser	Gly	Pro	Lys	Leu	Ser	Ser	Pro	Val	Thr
145				150						155					160
Gln	Asn	Val	Thr	Phe	Val	Asn	Gln	Gly	Gly	Pro	Met	Leu	Ser	Arg	Ala
			165						170						175
Ser	Ala	Met	Glu	Leu	Leu	Gln	Ala	Ala	Gly	Xaa					
	180							185							

2 INFORMATION FOR SEQ ID NO:54

REFERENCE CHARACTERISTICS

- A LENGTH 148 amino acids
B TYPE amino acid
C STRANDEDNESS single
D TOPOLOGY linear

XL. SEQUENCE DESCRIPTION: SEQ ID NO. 64

[illegible]

```

      50              55              60
Ala Gly Arg His Pro Asp Ser Asp Ile Phe Leu Asp Asp Val Thr Val
65              70              75              80
Ser Arg Arg His Ala Glu Phe Arg Leu Glu Asn Asn Glu Phe Asn Val
      85              90              95
Val Asp Val Gly Ser Leu Asn Gly Thr Tyr Val Asn Arg Glu Pro Val
      100             105             110
Asp Ser Ala Val Leu Ala Asn Gly Asp Glu Val Gln Ile Gly Lys Leu
      115             120             125
Arg Leu Val Phe Leu Thr Gly Pro Lys Gln Gly Glu Asp Asp Gly Ser
      130             135             140
Thr Gly Gly Pro
145

```

(2) INFORMATION FOR SEQ ID NO:65:

```

(1) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 230 amino acids
      (B) TYPE: amino acid
      (C) STRANDEDNESS: single
      (D) TOPOLOGY: linear

```

(ix) SEQUENCE DESCRIPTION: SEQ ID NO:65:

```

Thr Ser Asn Arg Pro Ala Arg Arg Gly Arg Arg Ala Pro Arg Asp Thr
1      5      10      15
Gly Pro Asp Arg Ser Ala Ser Leu Ser Leu Val Arg His Arg Arg Gln
      20      25      30
Gln Arg Asp Ala Leu Cys Leu Ser Thr Gln Ile Ser Arg Gln Ser
      35      40      45
Asn Leu Pro Pro Ala Ala Gly Gly Ala Ala Asn Tyr Ser Arg Arg Asn
      50      55      60
Phe Asp Val Arg Ile Lys Ile Phe Met Leu Val Thr Ala Val Val Leu
      65      70      75      80
Leu Cys Cys Ser Gly Val Ala Thr Ala Ala Pro Lys Thr Tyr Cys Glu
      85      90      95
Glu Leu Lys Gly Thr Asp Thr Gly Gln Ala Cys Gln Ile Gln Met Ser
      100     105     110
Asp Pro Ala Tyr Asn Ile Asn Ile Ser Leu Pro Ser Tyr Tyr Pro Asp
      115     120     125
Gln Lys Ser Leu Thr Asn Tyr Ile Ala Gln Thr Arg Asp Lys Phe Leu
      130     135     140
Ser Ala Ala Thr Ser Ser Thr Pro Arg Glu Ala Pro Tyr Glu Leu Asn
      145     150     155
Ile Thr Ser Ala Thr Tyr Gln Ser Ala Ile Pro Pro Arg Gly Thr Gln
      160     165     170     175
Ala Val Val Leu Asn Val Tyr His Asn Ala Gly Gly Thr His Pro Thr
      180     185     190
Thr Thr Thr Lys Ala Phe Asp Pro Asp Gln Ala Thr Arg Lys Pro Ile
      195     200     205
Thr Thr Asp Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr

```

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 132 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

```

Thr Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gly Gln Gly Phe
1      5      10      15
Ala Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser
20      25      30
Gly Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly
35      40      45
Leu Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val
50      55      60
Val Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val
65      70      75      80
Ile Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala
85      90      95
Asp Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Asn Trp
100     105     110
Gln Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Glu
115     120     125
Gly Pro Pro Ala
130

```

2. INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 100 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

```

Val Pro Leu Arg Ser Pro Ser Met Ser Pro Ser Lys Cys Leu Ala Ala
1      5      10      15
Ala His Arg Asn Pro Val Ile Arg Arg Arg Arg Leu Ser Asn Pro Pro
20      25      30
Pro Arg Lys Tyr Arg Ser Met Val Ser Pro Ala Thr Ala Ser Ala Gly
35      40      45
Met Ala Arg Val Arg Arg Arg Ala Ile Trp Arg Gly Pro Ala Thr Xaa
50      55      60
Ser Ala Gly Met Ala Arg Val Arg Arg Trp Xaa Val Met Pro Xaa Val
65      70      75      80
Ile Gln Ser Thr Xaa Ile Arg Xaa Xaa Gly Pro Ser Asp Asn Arg Gly
85      90      95
Ser Ala Arg Lys
100

```


(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 163 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

```

Met Thr Asp Asp Ile Leu Leu Ile Asp Thr Asp Glu Arg Val Arg Thr
1           5           10           15
Leu Thr Leu Asn Arg Pro Gln Ser Arg Asn Ala Leu Ser Ala Ala Leu
20           25           30
Arg Asp Arg Phe Phe Ala Xaa Leu Xaa Asp Ala Glu Xaa Asp Asp Asp
35           40           45
Ile Asp Val Val Ile Leu Thr Gly Ala Asp Pro Val Phe Cys Ala Gly
50           55           60
Leu Asp Leu Lys Val Ala Gly Arg Ala Asp Arg Ala Ala Gly His Leu
65           70           75           80
Thr Ala Val Gly Gly His Asp Gln Ala Gly Asp Arg Arg Asp Gln Arg
85           90           95
Arg Arg Gly His Arg Arg Ala Arg Thr Gly Ala Val Leu Arg His Pro
100          105          110
Asp Arg Leu Arg Ala Arg Pro Leu Arg Arg His Pro Arg Pro Gly Gly
115          120          125
Ala Ala Ala His Leu Gly Thr Gln Cys Val Leu Ala Ala Lys Gly Arg
130          135          140
His Arg Xaa Gly Pro Val Asp Glu Pro Asp Arg Arg Leu Pro Val Arg
145          150          155          160
Asp Arg Arg

```

(ii) INFORMATION FOR SEQ ID NO:69:

1. SEQUENCE CHARACTERISTICS

- (A) LENGTH: 344 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

xi. SEQUENCE DESCRIPTION: SEQ ID NO:69:

```

Met Lys Phe Val Asn His Ile Glu Pro Val Ala Pro Arg Arg Ala Gly
1           5           10           15
Arg Ala Val Ala Glu Val Tyr Ala Glu Ala Arg Arg Glu Phe Gly Arg
20           25           30
Leu Pro Glu Pro Leu Ala Met Leu Ser Pro Asp Glu Gly Leu Leu Thr
35           40           45
Ala Gly Trp Ala Thr Leu Arg Glu Thr Leu Leu Val Gly Gln Val Pro
50           55           60
Arg Gly Arg Lys Glu Ala Val Ala Ala Ala Thr Ala Ala Ser Leu Arg
65           70           75           80

```

100

Gly Asp	Pro Asn	Ala Pro	Tyr Val	Ala Trp	Ala Ala	Gly Thr	Gly Thr		
115			120			125			
Pro Ala	Gly Pro	Pro Ala	Pro Phe	Gly Pro	Asp Val	Ala Ala	Glu Tyr		
130			135			140			
Leu Gly	Thr Ala	Val Gln	Phe His	Phe Ile	Ala Arg	Leu Val	Leu Val		
145		150			155				
Leu Leu	Asp Glu	Thr Phe	Leu Pro	Gly Gly	Pro Arg	Ala Gln	Gln Leu		
		165			170				
Met Arg	Arg Ala	Gly Gly	Leu Val	Phe Ala	Arg Lys	Val Arg	Ala Glu		
		180			185				
His Arg	Pro Gly	Arg Ser	Thr Arg	Arg Leu	Glu Pro	Arg Thr	Leu Pro		
		195		200		205			
Asp Asp	Leu Ala	Trp Ala	Thr Pro	Ser Glu	Pro Ile	Ala Thr	Ala Phe		
		210		215		220			
Ala Ala	Leu Ser	His His	Leu Asp	Thr Ala	Pro His	Leu Pro	Pro Pro		
225		230			235				
Thr Arg	Gln Val	Val Arg	Arg Val	Val Gly	Ser Trp	His Gly	Glu Phe		
		245			250				
Met Pro	Met Ser	Ser Arg	Trp Thr	Asn Glu	His Thr	Ala Glu	Leu Pro		
		260		265		270			
Ala Asp	Leu His	Ala Pro	Thr Arg	Leu Ala	Leu Leu	Thr Gly	Leu Ala		
		275		280		285			
Pro His	Gln Val	Thr Asp	Asp Asp	Val Ala	Ala Ala	Ala Arg	Ser Leu	Leu	
		290		295		300			
Asp Thr	Asp Ala	Ala Leu	Val Gly	Ala Leu	Ala Trp	Ala Ala	Phe Thr		
305		310			315				
Ala Ala	Arg Arg	Ile Gly	Thr Trp	Ile Gly	Ala Ala	Ala Glu	Gly Gln		
		325			330				
Val Ser	Arg Gln	Asn Pro	Thr Gly						
		340							

2: INFORMATION FOR SEQ ID NO: 70:

2. SEQUENCE CHARACTERISTICS

- (A) LENGTH: 485 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

SI	SEQUENCE DESCRIPTION	SEQ ID NO
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

[illegible]

Asp Gln Tyr Glu Pro Gly Ser Ser Arg Arg Trp Ala Glu Arg Phe Ala
 115 120 125
 Thr Leu Leu Arg Asn Leu Glu Phe Leu Pro Asn Ser Pro Thr Leu Met
 130 135 140
 Asn Ser Gly Thr Asp Leu Gly Leu Leu Ala Gly Cys Phe Val Leu Pro
 145 150 155 160
 Ile Glu Asp Ser Leu Gln Ser Ile Phe Ala Thr Leu Gly Gln Ala Ala
 165 170 175
 Glu Leu Gln Arg Ala Gly Gly Gly Thr Gly Tyr Ala Phe Ser His Leu
 180 185 190
 Arg Pro Ala Gly Asp Arg Val Ala Ser Thr Gly Gly Thr Ala Ser Gly
 195 200 205
 Pro Val Ser Phe Leu Arg Leu Tyr Asp Ser Ala Ala Gly Val Val Ser
 210 215 220
 Met Gly Gly Arg Arg Arg Gly Ala Cys Met Ala Val Leu Asp Val Ser
 225 230 235 240
 His Pro Asp Ile Cys Asp Phe Val Thr Ala Lys Ala Glu Ser Pro Ser
 245 250 255
 Glu Leu Pro His Phe Asn Leu Ser Val Gly Val Thr Asp Ala Phe Leu
 260 265 270
 Arg Ala Val Glu Arg Asn Gly Leu His Arg Leu Val Asn Pro Arg Thr
 275 280 285
 Gly Lys Ile Val Ala Arg Met Pro Ala Ala Glu Phe Asp Ala Ile
 290 295 300
 Cys Lys Ala Ala His Ala Gly Gly Asp Pro Gly Leu Val Phe Leu Asp
 305 310 315 320
 Thr Ile Asn Arg Ala Asn Pro Val Pro Gly Arg Gly Arg Ile Glu Ala
 325 330 335
 Thr Asn Pro Cys Gly Glu Val Pro Leu Leu Pro Tyr Glu Ser Cys Asn
 340 345 350
 Leu Gly Ser Ile Asn Leu Ala Arg Met Leu Ala Asp Gly Arg Val Asp
 355 360 365
 Trp Asp Arg Leu Gln His Val Ala Gly Val Ala Val Arg Phe Leu Asp
 370 375 380
 Asp Val Ile Asp Val Ser Arg Tyr Pro Phe Pro Glu Leu Gly Glu Ala
 385 390 395 400
 Ala Arg Ala Thr Arg Lys Ile Gly Leu Gly Val Met Gly Leu Ala Glu
 405 410 415
 Leu Leu Ala Ala Leu Gly Ile Pro Tyr Asp Ser Glu Glu Ala Val Arg
 420 425 430
 Leu Ala Thr Arg Leu Met Arg Arg Ile Gln Gln Ala Ala His Thr Ala
 435 440 445
 Ser Arg Arg Leu Ala Gln Gln Arg Gly Ala Phe Pro Ala Phe Thr Asp
 450 455 460
 Ser Arg Phe Ala Arg Ser Gly Pro Arg Arg Asn Ala His Val Thr Ser
 465 470 475 480
 Val Ala Pro Thr Gly
 485

2 INFORMATION FOR SEQ ID NO 21:

SEQUENCE INFORMATION

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

Gly Val Ile Val Leu Asp Leu Glu Pro Arg Gly Pro Leu Pro Thr Glu
 1 5 10 15
 Ile Tyr Trp Arg Arg Arg Gly Leu Ala Leu Gly Ile Ala Val Val Val
 20 25 30
 Val Gly Ile Ala Val Ala Ile Val Ile Ala Phe Val Asp Ser Ser Ala
 35 40 45
 Gly Ala Lys Pro Val Ser Ala Asp Lys Pro Ala Ser Ala Gln Ser His
 50 55 60
 Pro Gly Ser Pro Ala Pro Gln Ala Pro Gln Pro Ala Gly Gln Thr Glu
 65 70 75 80
 Gly Asn Ala Ala Ala Pro Pro Gln Gly Gln Asn Pro Glu Thr Pro
 85 90 95
 Thr Pro Thr Ala Ala Val Gln Pro Pro Val Leu Lys Glu Gly Asp
 100 105 110
 Asp Cys Pro Asp Ser Thr Leu Ala Val Lys Gly Leu Thr Asn Ala Pro
 115 120 125
 Gln Tyr Tyr Val Gly Asp Gln Pro Lys Phe Thr Met Val Val Thr Asn
 130 135 140
 Ile Gly Leu Val Ser Cys Lys Arg Asp Val Gly Ala Ala Val Leu Ala
 145 150 155 160
 Ala Tyr Val Tyr Ser Leu Asp Asn Lys Arg Leu Trp Ser Asn Leu Asp
 165 170 175
 Cys Ala Pro Ser Asn Gln Thr Leu Val Lys Thr Phe Ser Pro Gly Glu
 180 185 190
 Gln Val Thr Thr Ala Val Thr Trp Thr Gly Met Gly Ser Ala Pro Arg
 195 200 205
 Cys Pro Leu Pro Arg Pro Ala Ile Gly Pro Gly Thr Tyr Asn Leu Val
 210 215 220
 Val Gln Leu Gly Asn Leu Ala Ser Leu Pro Val Pro Phe Ile Leu Asn
 225 230 235 240
 Gln Pro Pro Pro Pro Gly Pro Val Pro Ala Pro Gly Pro Ala Gln
 245 250 255
 Ala Pro Pro Pro Glu Ser Pro Ala Gln Gly Gly
 260 265

1 INFORMATION FOR SEQ ID NO:72:

- 1 SEQUENCE CHARACTERISTICS
 A LENGTH: 265 amino acids
 B TYPE: amino acid
 C STRANDEDNESS: single
 D TOPOLOGY: linear

xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

Leu Ile Ser Thr Tyr Lys Ala Ser His Ala Ser Leu Gly Val Gln Val
 1

Lys Val Asp Asp Arg Pro Ile Asn Ser Ala Asp Ala Leu Val Ala Ala
 50 55 60
 Val Arg Ser Lys Ala Pro Gly Ala Thr Val Ala Leu Thr Phe Gln Asp
 65 70 75 80
 Pro Ser Gly Gly Ser Arg Thr Val Gln Val Thr Leu Gly Lys Ala Glu
 85 90 95
 Gln

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 364 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

Gly Ala Ala Val Ser Leu Leu Ala Ala Gly Thr Leu Val Leu Thr Ala
 1 5 10 15
 Cys Gly Gly Gly Thr Asn Ser Ser Ser Ser Gly Ala Gly Gly Thr Ser
 20 25 30
 Gly Ser Val His Cys Gly Gly Lys Lys Glu Leu His Ser Ser Gly Ser
 35 40 45
 Thr Ala Gln Glu Asn Ala Met Gln Gln Phe Val Tyr Ala Tyr Val Arg
 50 55 60
 Ser Cys Pro Gly Tyr Thr Leu Asp Tyr Asn Ala Asn Gly Ser Gly Ala
 65 70 75 80
 Gly Val Thr Gln Phe Leu Asn Asn Gln Thr Asp Phe Ala Gly Ser Asp
 85 90 95
 Val Pro Leu Asn Pro Ser Thr Gly Gln Pro Asp Arg Ser Ala Glu Arg
 100 105 110
 Cys Gly Ser Pro Ala Trp Asp Leu Pro Thr Val Phe Gly Pro Ile Ala
 115 120 125
 Ile Thr Tyr Asn Ile Lys Gly Val Ser Thr Leu Asn Leu Asp Gly Pro
 130 135 140
 Thr Thr Ala Lys Ile Phe Asn Gly Thr Ile Thr Val Trp Asn Asp Pro
 145 150 155 160
 Gln Ile Gln Ala Leu Asn Ser Gly Thr Asn Leu Pro Pro Thr Pro Ile
 165 170 175
 Ser Val Ile Phe Arg Ser Asp Lys Ser Gly Thr Ser Asp Asn Phe Gln
 180 185 190
 Lys Thr Leu Asp Ile Val Ser Asn Gly Ala Trp Gly Lys Gly Ala Ser
 195 200 205
 Ala Thr Phe Ser Gly Gly Val Gly Val Gly Ala Ser Gly Asn Asn Gly
 210 215 220
 Thr Ser Ala Leu Leu Gln Thr Thr Asp Gly Ser Ile Thr Tyr Asn Glu
 225 230 235 240
 Trp Ser Phe Ala Val Gly Lys Gln Leu Asn Met Ala Gln Ile Ile Thr
 245 250 255
 Ser Ala Gly Trp Asn Ile Val Ala Thr Thr Thr Thr Thr Thr Thr Thr

```

      290              295              300
Val Leu Ala Thr Tyr Glu Ile Val Cys Ser Lys Tyr Pro Asp Ala Thr
305              310              315              320
Thr Gly Thr Ala Val Arg Ala Phe Met Gln Ala Ala Ile Gly Pro Gly
      325              330              335
Gln Glu Gly Leu Asp Gln Tyr Gly Ser Ile Pro Leu Pro Lys Ser Phe
      340              345              350
Gln Ala Lys Leu Ala Ala Ala Val Asn Ala Ile Ser
      355              360

```

(2) INFORMATION FOR SEQ ID NO:74:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 309 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:74:

```

Gln Ala Ala Ala Gly Arg Ala Val Arg Arg Thr Gly His Ala Glu Asp
1           5           10           15
Gln Thr His Gln Asp Arg Leu His His Gly Cys Arg Arg Ala Ala Val
20           25           30
Val Val Arg Gln Asp Arg Ala Ser Val Ser Ala Thr Ser Ala Arg Pro
35           40           45
Pro Arg Arg His Pro Ala Gln Gly His Arg Arg Arg Val Ala Pro Ser
50           55           60
Gly Gly Arg Arg Arg Pro His Pro His His Val Gln Pro Asp Asp Arg
65           70           75           80
Arg Asp Arg Pro Ala Leu Leu Asp Arg Thr Gln Pro Ala Glu His Pro
85           90           95
Asp Pro His Arg Arg Gly Pro Ala Asp Pro Gly Arg Val Arg Gly Arg
100          105          110
Gly Arg Leu Arg Arg Val Asp Asp Gly Arg Leu Gln Pro Asp Arg Asp
115          120          125
Ala Asp His Gly Ala Pro Val Arg Gly Arg Gly Pro His Arg Gly Val
130          135          140
Gln His Arg Gly Gly Pro Val Phe Val Arg Arg Val Pro Gly Val Arg
145          150          155          160
Cys Ala His Arg Arg Gly His Arg Arg Val Ala Ala Pro Gly Gln Gly
165          170          175
Asp Val Leu Arg Ala Gly Leu Arg Val Gln Arg Leu Arg Pro Val Ala
180          185          190
Ala Val Ala Asn Leu His Arg Gly Ser Gln Arg Ala Asp Gly Arg Val
195          200          205
Phe Arg Pro Ile Arg Arg Gly Ala Arg Leu Pro Ala Arg Arg Ser Arg
210          215          220
Ala Gly Pro Gln Gly Arg Leu His Leu Asn Gly Ala Gly Pro Ser Pro
225          230          235          240
Leu Ser Ala Arg Ala Gly Gln Gln Arg Arg Arg Arg Arg Arg Arg Arg

```

275
Ala Gly Val Ala His Ala Ala Gly Pro Arg Arg 285
290 295 300
Asn Arg Pro Arg Arg
305

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 580 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

Ser	Ala	Val	Trp	Cys	Leu	Asn	Gly	Phe	Thr	Gly	Arg	His	Arg	His	Gly
1									10					15	
Arg	Cys	Arg	Val	Arg	Ala	Ser	Gly	Trp	Arg	Ser	Ser	Asn	Arg	Trp	Cys
		20						25					30		
Ser	Thr	Thr	Ala	Asp	Cys	Cys	Ala	Ser	Lys	Thr	Pro	Thr	Gln	Ala	Ala
		35					40					45			
Ser	Pro	Leu	Glu	Arg	Arg	Phe	Thr	Cys	Cys	Ser	Pro	Ala	Val	Gly	Cys
		50				55					60				
Arg	Phe	Arg	Ser	Phe	Pro	Val	Arg	Arg	Leu	Ala	Leu	Gly	Ala	Arg	Thr
65					70					75				80	
Ser	Arg	Thr	Leu	Gly	Val	Arg	Arg	Thr	Leu	Ser	Gln	Trp	Asn	Leu	Ser
			85						90					95	
Pro	Arg	Ala	Gln	Pro	Ser	Cys	Ala	Val	Thr	Val	Glu	Ser	His	Thr	His
		100						105					110		
Ala	Ser	Pro	Arg	Met	Ala	Lys	Leu	Ala	Arg	Val	Val	Gly	Leu	Val	Gln
		115					120						125		
Glu	Gln	Gln	Pro	Ser	Asp	Met	Thr	Asn	His	Pro	Arg	Tyr	Ser	Pro	Pro
130						135						140			
Pro	Gln	Gln	Pro	Gly	Thr	Pro	Gly	Tyr	Ala	Gln	Gly	Gln	Gln	Gln	Thr
145					150					155				160	
Tyr	Ser	Gln	Gln	Phe	Asp	Trp	Arg	Tyr	Pro	Pro	Ser	Pro	Pro	Pro	Gln
				165					170					175	
Pro	Thr	Gln	Tyr	Arg	Gln	Pro	Tyr	Gln	Ala	Leu	Gly	Gly	Thr	Arg	Pro
		180						185					190		
Gly	Leu	Ile	Pro	Gly	Val	Ile	Pro	Thr	Met	Thr	Pro	Pro	Pro	Gly	Met
195							200					205			
Val	Arg	Gln	Arg	Pro	Arg	Ala	Gly	Met	Leu	Ala	Ile	Gly	Ala	Val	Thr
	210					215					220				
Ile	Ala	Val	Val	Ser	Ala	Gly	Ile	Gly	Gly	Ala	Ala	Ala	Ser	Leu	Val
225					230					235				240	
Gly	Phe	Asn	Arg	Ala	Pro	Ala	Gly	Pro	Ser	Gly	Gly	Pro	Val	Ala	Ala
				245				250					255		
Ser	Ala	Ala	Pro	Ser	Ile	Pro	Ala	Ala	Asn	Met	Pro	Pro	Gly	Ser	Val
		260						265					270		
Gln	Gln	Val	Ala	Ala	Leu	Val	Trp	Val	Val	Val	Val	Val	Val	Val	Val

305 310 315 320
 Pro Pro Leu Gly Ser Pro Pro Pro Lys Thr Thr Val Thr Phe Ser Asp
 325 330 335
 Gly Arg Thr Ala Pro Phe Thr Val Val Gly Ala Asp Pro Thr Ser Asp
 340 345 350
 Ile Ala Val Val Arg Val Gln Gly Val Ser Gly Leu Thr Pro Ile Ser
 355 360 365
 Leu Gly Ser Ser Ser Asp Leu Arg Val Gly Gln Pro Val Leu Ala Ile
 370 375 380
 Gly Ser Pro Leu Gly Leu Glu Gly Thr Val Thr Thr Gly Ile Val Ser
 385 390 395 400
 Ala Leu Asn Arg Pro Val Ser Thr Thr Gly Glu Ala Gly Asn Gln Asn
 405 410 415
 Thr Val Leu Asp Ala Ile Gln Thr Asp Ala Ala Ile Asn Pro Gly Asn
 420 425 430
 Ser Gly Gly Ala Leu Val Asn Met Asn Ala Gln Leu Val Gly Val Asn
 435 440 445
 Ser Ala Ile Ala Thr Leu Gly Ala Asp Ser Ala Asp Ala Gln Ser Gly
 450 455 460
 Ser Ile Gly Leu Gly Phe Ala Ile Pro Val Asp Gln Ala Lys Arg Ile
 465 470 475 480
 Ala Asp Glu Leu Ile Ser Thr Gly Lys Ala Ser His Ala Ser Leu Gly
 485 490 495
 Val Gln Val Thr Asn Asp Lys Asp Thr Pro Gly Ala Lys Ile Val Glu
 500 505 510
 Val Val Ala Gly Gly Ala Ala Ala Asn Ala Gly Val Pro Lys Gly Val
 515 520 525
 Val Val Thr Lys Val Asp Asp Arg Pro Ile Asn Ser Ala Asp Ala Leu
 530 535 540
 Val Ala Ala Val Arg Ser Lys Ala Pro Gly Ala Thr Val Ala Leu Thr
 545 550 555 560
 Phe Gln Asp Pro Ser Gly Gly Ser Arg Thr Val Gln Val Thr Leu Gly
 565 570 575
 Lys Ala Glu Gln
 580

1. INFORMATION FOR SEQ ID NO:76:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 580 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:76

Met Asn Asp Gly Lys Arg Ala Val Thr Ser Ala Val Leu Val Val Leu
 1 5 10 15
 Gly Ala Cys Leu Ala Leu Arg Leu Ser Gly Cys Ser Ser Pro Lys Pro
 20 25 30
 Asp Ala Ala Gln Gln Val Pro Val Ser Pro Thr Ala Ser Ser Pro
 35 40 45


```

65          70          75          80
Leu Gly Ile Thr Ser Ala Asp Val Asp Val Arg Ala Asn Pro Leu Ala
          85          90          95
Ala Lys Gly Val Cys Thr Tyr Asn Asp Glu Gln Gly Val Pro Phe Arg
          100          105          110
Val Gln Gly Asp Asn Ile Ser Val Lys Leu Phe Asp Asp Trp Ser Asn
          115          120          125
Leu Gly Ser Ile Ser Glu Leu Ser Thr Ser Arg Val Leu Asp Pro Ala
          130          135          140
Ala Gly Val Thr Gln Leu Leu Ser Gly Val Thr Asn Leu Gln Ala Gln
145          150          155          160
Gly Thr Glu Val Ile Asp Gly Ile Ser Thr Thr Lys Ile Thr Gly Thr
          165          170          175
Ile Pro Ala Ser Ser Val Lys Met Leu Asp Pro Gly Ala Lys Ser Ala
          180          185          190
Arg Pro Ala Thr Val Trp Ile Ala Gln Asp Gly Ser His His Leu Val
          195          200          205
Arg Ala Ser Ile Asp Leu Gly Ser Gly Ser Ile Gln Leu Thr Gln Ser
          210          215          220
Lys Trp Asn Glu Pro Val Asn Val Asp
225          230

```

(2) INFORMATION FOR SEQ ID NO:77:

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 66 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:77:

```

Val Ile Asp Ile Ile Gly Thr Ser Pro Thr Ser Trp Gln Gln Ala Ala
1      5      10      15
Ala Gln Ala Val Gln Arg Ala Arg Asp Ser Val Asp Asp Ile Arg Val
20      25      30
Ala Arg Val Ile Glu Gln Asp Met Ala Val Asp Ser Ala Gly Lys Ile
35      40      45
Thr Tyr Arg Ile Lys Leu Gln Val Ser Phe Lys Met Arg Pro Ala Gln
50      55      60
Pro Arg
65

```

(3) INFORMATION FOR SEQ ID NO:78:

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 69 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO:78:

Cys Ala Ser Pro Pro Ser Pro Pro Leu Pro Pro Ala Pro Pro Val Ala
20 25 30
Pro Gly Pro Pro Met Pro Pro Leu Asp Pro Trp Pro Pro Ala Pro Pro
35 40 45
Leu Pro Tyr Ser Thr Pro Pro Gly Ala Pro Leu Pro Pro Ser Pro Pro
50 55 60
Ser Pro Pro Leu Pro
65

(2) INFORMATION FOR SEQ ID NO. 79:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 355 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:70.

Met	Ser	Asn	Ser	Arg	Arg	Arg	Ser	Leu	Arg	Trp	Ser	Trp	Leu	Leu	Ser
1				5					10					15	
Val	Leu	Ala	Ala	Val	Gly	Leu	Gly	Leu	Ala	Thr	Ala	Pro	Ala	Gln	Ala
		20						25					30		
Ala	Pro	Pro	Ala	Leu	Ser	Gln	Asp	Arg	Phe	Ala	Asp	Phe	Pro	Ala	Leu
		35					40					45			
Pro	Leu	Asp	Pro	Ser	Ala	Met	Val	Ala	Gln	Val	Ala	Pro	Gln	Val	Val
		50				55					60				
Asn	Ile	Asn	Thr	Lys	Leu	Gly	Tyr	Asn	Asn	Ala	Val	Gly	Ala	Gly	Thr
65				70						75				80	
Gly	Ile	Val	Ile	Asp	Pro	Asn	Gly	Val	Val	Leu	Thr	Asn	Asn	His	Val
				85					90				95		
Ile	Ala	Gly	Ala	Thr	Asp	Ile	Asn	Ala	Phe	Ser	Val	Gly	Ser	Gly	Gln
		100					105					110			
Thr	Tyr	Gly	Val	Asp	Val	Val	Gly	Tyr	Asp	Arg	Thr	Gln	Asp	Val	Ala
		115					120					125			
Val	Leu	Gln	Leu	Arg	Gly	Ala	Gly	Gly	Leu	Pro	Ser	Ala	Ala	Ile	Gly
		130			135					140					
Gly	Gly	Val	Ala	Val	Gly	Gln	Pro	Val	Val	Ala	Met	Gly	Asn	Ser	Gly
145					150					155				160	
Gly	Gln	Gly	Gly	Thr	Pro	Arg	Ala	Val	Pro	Gly	Arg	Val	Val	Ala	Leu
				165					170				175		
Gly	Gln	Thr	Val	Gln	Ala	Ser	Asp	Ser	Leu	Thr	Gly	Ala	Gln	Glu	Thr
			180				185					190			
Leu	Asn	Gly	Leu	Ile	Gln	Phe	Asp	Ala	Ala	Ile	Gln	Pro	Gly	Asp	Ser
		195				200						205			
Gly	Gly	Pro	Val	Val	Asn	Gly	Leu	Gly	Gln	Val	Val	Gly	Met	Asn	Thr
		210				215						220			
Ala	Ala	Ser	Asp	Asn	Phe	Gln	Leu	Ser	Gln	Gly	Gln	Gln	Gly	Phe	Ala
225					230						235				
Ile	Pro	Ile	Gly	Gln	Ala	Met	Ala	Ile	Ala	Gly	Gln	Ile	Arg	Ser	Gly
				240					245						
His	His	Ser	Asn	Met					250						

Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val Ile
 290 295 300
 Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala Asp
 305 310 315 320
 Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Asn Trp Gln
 325 330 335
 Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Glu Gly
 340 345 350
 Pro Pro Ala
 355

(2) INFORMATION FOR SEQ ID NO:80:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 205 amino acids
- (B) TYPE: amino acid
- (C) STANDEDNESS: single
- (D) TOPOLOGY: linear

(11) SEQUENCE DESCRIPTION: SEQ ID NO:80:

Ser Pro Lys Pro Asp Ala Glu Glu Gln Gly Val Pro Val Ser Pro Thr
 1 5 10 15
 Ala Ser Asp Pro Ala Leu Leu Ala Glu Ile Arg Gln Ser Leu Asp Ala
 20 25 30
 Thr Lys Gly Leu Thr Ser Val His Val Ala Val Arg Thr Thr Gly Lys
 35 40 45
 Val Asp Ser Leu Leu Gly Ile Thr Ser Ala Asp Val Asp Val Arg Ala
 50 55 60
 Asn Pro Leu Ala Ala Lys Gly Val Cys Thr Tyr Asn Asp Glu Gln Gly
 65 70 75 80
 Val Pro Phe Arg Val Gln Gly Asp Asn Ile Ser Val Lys Leu Phe Asp
 85 90 95
 Asp Trp Ser Asn Leu Gly Ser Ile Ser Glu Leu Ser Thr Ser Arg Val
 100 105 110
 Leu Asp Pro Ala Ala Gly Val Thr Gln Leu Leu Ser Gly Val Thr Asn
 115 120 125
 Leu Gln Ala Gln Gly Thr Glu Val Ile Asp Gly Ile Ser Thr Thr Lys
 130 135 140
 Ile Thr Gly Thr Ile Pro Ala Ser Ser Val Lys Met Leu Asp Pro Gly
 145 150 155 160
 Ala Lys Ser Ala Arg Pro Ala Thr Val Trp Ile Ala Gln Asp Gly Ser
 165 170 175
 His His Leu Val Arg Ala Ser Ile Asp Leu Gly Ser Gly Ser Ile Gln
 180 185 190
 Leu Thr Gln Ser Cys Trp Asn Gln Pro Val Asn Val Asp
 195 200 205

(2) INFORMATION FOR SEQ ID NO:81:

(1) SEQUENCE CHARACTERISTICS:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:81:

Gly Asp Ser Phe Trp Ala Ala Ala Asp Gln Met Ala Arg Gly Phe Val
 1 5 10 15
 Leu Gly Ala Thr Ala Gly Arg Thr Thr Leu Thr Gly Glu Gly Leu Gln
 20 25 30
 His Ala Asp Gly His Ser Leu Leu Leu Asp Ala Thr Asn Pro Ala Val
 35 40 45
 Val Ala Tyr Asp Pro Ala Phe Ala Tyr Glu Ile Gly Tyr Ile Xaa Glu
 50 55 60
 Ser Gly Leu Ala Arg Met Cys Gly Glu Asn Pro Glu Asn Ile Phe Phe
 65 70 75 80
 Tyr Ile Thr Val Tyr Asn Glu Pro Tyr Val Gln Pro Pro Glu Pro Glu
 85 90 95
 Asn Phe Asp Pro Glu Gly Val Leu Gly Gly Ile Tyr Arg Tyr His Ala
 100 105 110
 Ala Thr Glu Gln Arg Thr Asn Lys Xaa Gln Ile Leu Ala Ser Gly Val
 115 120 125
 Ala Met Pro Ala Ala Leu Arg Ala Ala Gln Met Leu Ala Ala Glu Trp
 130 135 140
 Asp Val Ala Ala Asp Val Trp Ser Val Thr Ser Trp Gly Glu Leu Asn
 145 150 155 160
 Arg Asp Gly Val Val Ile Glu Thr Glu Lys Leu Arg His Pro Asp Arg
 165 170 175
 Pro Ala Gly Val Pro Tyr Val Thr Arg Ala Leu Glu Asn Ala Arg Gly
 180 185 190
 Pro Val Ile Ala Val Ser Asp Trp Met Arg Ala Val Pro Glu Gln Ile
 195 200 205
 Arg Pro Trp Val Pro Gly Thr Tyr Leu Thr Leu Gly Thr Asp Gly Phe
 210 215 220
 Gly Phe Ser Asp Thr Arg Pro Ala Gly Arg Arg Tyr Phe Asn Thr Asp
 225 230 235 240
 Ala Gln Ser Gln Val Gly Arg Gly Phe Gly Arg Gly Trp Pro Gly Arg
 245 250 255
 Arg Val Asn Ile Asp Pro Phe Gly Ala Gly Arg Gly Pro Pro Ala Gln
 260 265 270
 Leu Pro Gly Phe Asp Gln Gly Gly Gly Leu Arg Pro Xaa Lys
 275 280 285

INFORMATION FOR SEQ ID NO: 81:

1. SEQUENCE CHARACTERISTICS:
 A. LENGTH: 285 amino acids
 B. TYPE: amino acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

X1. SEQUENCE DESCRIPTION: SEQ ID NO:81:

Phe Lys Phe His Ala Leu Met Gln Thr Gln Ile His Asn Glu Asp Thr

111

```

      35              40              45
Asn His Ala Met Met Leu Val Gln His Leu Leu Asp Arg Asp Leu Arg
 50              55              60
Val Glu Ile Pro Gly Val Asp Thr Val Arg Asn Gln Phe Asp Arg Pro
 65              70              75              80
Arg Glu Ala Leu Ala Leu Ala Leu Asp Gln Glu Arg Thr Val Thr Asp
      85              90              95
Gln Val Gly Arg Leu Thr Ala Val Ala Arg Asp Glu Gly Asp Phe Leu
 100              105              110
Gly Glu Gln Phe Met Gln Trp Phe Leu Gln Glu Gln Ile Glu Glu Val
 115              120              125
Ala Leu Met Ala Thr Leu Val Arg Val Ala Asp Arg Ala Gly Ala Asn
 130              135              140
Leu Phe Glu Leu Glu Asn Phe Val Ala Arg Glu Val Asp Val Ala Pro
 145              150              155              160
Ala Ala Ser Gly Ala Pro His Ala Ala Gly Gly Arg Leu
      165              170

```

2. INFORMATION FOR SEQ ID NO:83:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 107 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:83:

```

Arg Ala Asp Glu Arg Lys Asn Thr Thr Met Lys Met Val Lys Ser Ile
 1              5              10              15
Ala Ala Gly Leu Thr Ala Ala Ala Ala Ile Gly Ala Ala Ala Gly
 20              25              30
Val Thr Ser Ile Met Ala Gly Gly Pro Val Val Tyr Gln Met Gln Pro
 35              40              45
Val Val Phe Gly Ala Pro Leu Pro Leu Asp Pro Xaa Ser Ala Pro Xaa
 50              55              60
Val Pro Thr Ala Ala Gln Trp Thr Xaa Leu Leu Asn Xaa Leu Xaa Asp
 65              70              75              80
Pro Asn Val Ser Phe Xaa Asn Lys Gly Ser Leu Val Gln Gly Gly Ile
 85              90              95
His Gly Xaa Gln Gly Xaa Xaa Arg Arg Xaa Gln
 100              105

```

3. INFORMATION FOR SEQ ID NO:84:

1. SEQUENCE CHARACTERISTICS:
- A. LENGTH: 126 amino acids
 - B. TYPE: amino acid
 - C. STRANDEDNESS: single
 - D. TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:84:

```

      20      25      30
Arg Arg Ala Leu Glu Leu Gln Ala Pro Ser Val Val Xaa Arg Gln Gly
      35      40      45
Val Lys Glu Pro Leu Xaa Thr Gly Ile Lys Ala Ile Asp Ala Met Thr
      50      55      60
Pro Ile Gly Arg Gly Gln Arg Gln Leu Ile Ile Gly Asp Arg Lys Thr
      65      70      75      80
Gly Lys Asn Arg Arg Leu Cys Arg Thr Pro Ser Ser Asn Gln Arg Glu
      85      90      95
Glu Leu Gly Val Arg Trp Ile Pro Arg Ser Arg Cys Ala Cys Val Tyr
      100      105      110
Val Gly His Arg Ala Arg Arg Gly Thr Tyr His Arg Arg
      115      120      125

```

(2) INFORMATION FOR SEQ ID NO:85:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 117 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:85:

```

Cys Asp Ala Val Met Gly Phe Leu Gly Gly Ala Gly Pro Leu Ala Val
1      5      10      15
Val Asp Gln Gln Leu Val Thr Arg Val Pro Gln Gly Trp Ser Phe Ala
      20      25      30
Gln Ala Ala Ala Val Pro Val Val Phe Leu Thr Ala Trp Tyr Gly Leu
      35      40      45
Ala Asp Leu Ala Glu Ile Lys Ala Gly Glu Ser Val Leu Ile His Ala
      50      55      60
Gly Thr Gly Gly Val Gly Met Ala Ala Val Gln Leu Ala Arg Gln Trp
      65      70      75      80
Gly Val Glu Val Phe Val Thr Ala Ser Arg Gly Lys Trp Asp Thr Leu
      85      90      95
Arg Ala Xaa Xaa Phe Asp Asp Xaa Pro Tyr Arg Xaa Phe Pro His Xaa
      100      105      110
Arg Ser Ser Xaa Gly
      115

```

(2) INFORMATION FOR SEQ ID NO:86:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 108 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:86:

Met Thr Arg Pro ...

Thr Ala Pro Val Pro Asp Tyr Tyr Trp Cys Pro Gly Gln Pro Phe Asp
35 40 45
Pro Ala Trp Gly Pro Asn Trp Asp Pro Tyr Thr Cys His Asp Asp Phe
50 55 60
His Arg Asp Ser Asp Gly Pro Asp His Ser Arg Asp Tyr Pro Gly Pro
65 70 75 80
Ile Leu Glu Gly Pro Val Leu Asp Asp Pro Gly Ala Ala Pro Pro Pro
85 90 95
Pro Ala Ala Gly Gly Gly Ala
100

(2) INFORMATION FOR SEQ ID NO:87:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

EX1) SEQUENCE DESCRIPTION: SEQ ID NO:37:

[illegible]

7 INFORMATION FOR SEQ ID NO: 88.

2. SEQUENCE CHARACTERISTICS:

- A) LENGTH: 95 amino acids
 B) TYPE: amino acid
 C) STRANDEDNESS: single
 D) TOPOLOGY: linear

ALL SEQUENCE DESCRIPTIONS ARE IN SEQ ID NO. 43

Thr Asp Ala Ala Thr Leu Ala Gln Gln Ala Gly Asn Phe Glu Arg Ile
1 5 10 15
Ser Gly Asp Leu Lys Thr Gln Ile Asp Gln Val Glu Ser Thr Ala Gly
20 25 30
Ser Leu Gln Gly Gln Tyr Arg Gly Ala Ala Gly Thr Ala Ala Gln Ala
35 40 45
Ala Val Val Arg Phe Leu Gln Ala Ala Ser Thr Thr Thr

85

90

95

(2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 166 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

```

Met Thr Gln Ser Gln Thr Val Thr Val Asp Gln Gln Glu Ile Leu Asn
1           5           10           15
Arg Ala Asn Glu Val Glu Ala Pro Met Ala Asp Pro Pro Thr Asp Val
20           25           30
Pro Ile Thr Pro Cys Glu Leu Thr Xaa Xaa Lys Asn Ala Ala Gln Gln
35           40           45
Xaa Val Leu Ser Ala Asp Asn Met Arg Glu Tyr Leu Ala Ala Gly Ala
50           55           60
Lys Glu Arg Gln Arg Leu Ala Thr Ser Leu Arg Asn Ala Ala Lys Xaa
65           70           75           80
Tyr Gly Glu Val Asp Glu Glu Ala Ala Thr Ala Leu Asp Asn Asp Gly
85           90           95
Glu Gly Thr Val Gln Ala Glu Ser Ala Gly Ala Val Gly Gly Asp Ser
100          105          110
Ser Ala Glu Leu Thr Asp Thr Pro Arg Val Ala Thr Ala Gly Glu Pro
115          120          125
Asn Phe Met Asp Leu Lys Glu Ala Ala Arg Lys Leu Glu Thr Gly Asp
130          135          140
Gln Gly Ala Ser Leu Ala His Xaa Gly Asp Gly Tyr Asn Thr Xaa Thr
145          150          155          160
Leu Thr Leu Gln Gly Asp
165

```

(2) INFORMATION FOR SEQ ID NO:90:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

```

Arg Ala Gln Arg Met
1           5

```

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

Val Ala Trp Met Ser Val Thr Ala Gly Gln Ala Glu Leu Thr Ala Ala
 1 5 10 15
 Gln Val Arg Val Ala Ala Ala Tyr Glu Thr Ala Tyr Gly Leu Thr
 20 25 30
 Val Pro Pro Pro Val Ile Ala Glu Asn Arg Ala Glu Leu Met Ile Leu
 35 40 45
 Ile Ala Thr Asn Leu Leu Gly Gln Asn Thr Pro Ala Ile Ala Val Asn
 50 55 60
 Glu Ala Glu Tyr Gly Glu Met Trp Ala Gln Asp Ala Ala Ala Met Phe
 65 70 75 80
 Gly Tyr Ala Ala Ala Thr Ala Thr Ala Thr Ala Thr Leu Leu Pro Phe
 85 90 95
 Glu Glu Ala Pro Glu Met Thr Ser Ala Gly Gly Leu Leu Gln Gln Ala
 100 105 110
 Ala Ala Val Glu Glu Ala Ser Asp Thr Ala Ala Ala Asn Gln Leu Met
 115 120 125
 Asn Asn Val Pro Gln Ala Leu Lys Gln Leu Ala Gln Pro Thr Gln Gly
 130 135 140
 Thr Thr Pro Ser Ser Lys Leu Gly Gly Leu Trp Lys Thr Val Ser Pro
 145 150 155 160
 His Arg Ser Pro Ile Ser Asn Met Val Ser Met Ala Asn Asn His Met
 165 170 175
 Ser Met Thr Asn Ser Gly Val Ser Met Thr Asn Thr Leu Ser Ser Met
 180 185 190
 Leu Lys Gly Phe Ala Pro Ala Ala Ala Gln Ala Val Gln Thr Ala
 195 200 205
 Ala Gln Asn Gly Val Arg Ala Met Ser Ser Leu Gly Ser Ser Leu Gly
 210 215 220
 Ser Ser Gly Leu Gly Gly Gly Val Ala Ala Asn Leu Gly Arg Ala Ala
 225 230 235 240
 Ser Val Arg Tyr Gly His Arg Asp Gly Gly Lys Tyr Ala Xaa Ser Gly
 245 250 255
 Arg Arg Asn Gly Gly Pro Ala
 260

2 INFORMATION FOR SEQ ID NO:92:

1. SEQUENCE CHARACTERISTICS:
 A. LENGTH: 263 amino acids
 B. TYPE: amino acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

XI SEQUENCE DESCRIPTION: SEQ ID NO:92:

Met Thr Tyr Ser Pro Gly Asn Pro Gly Tyr Pro Gln Ala Gln Pro Ala
 1 10 15
 Gly Ser Thr Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser
 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260

50 55 60
 Glu Leu Gly Gly Gly Asp Gly Ala Val Ser Gly Asp Thr Gly Leu Pro
 65 70 75 80
 Val Gly Val Ala Leu Leu Ala Ala Leu Leu Ala Gly Val Val Leu Val
 85 90 95
 Pro Lys Ala Lys Ser His Val Thr Val Val Ala Val Leu Gly Val Leu
 100 105 110
 Gly Val Phe Leu Met Val Ser Ala Thr Phe Asn Lys Pro Ser Ala Tyr
 115 120 125
 Ser Thr Gly Trp Ala Leu Trp Val Val Leu Ala Phe Ile Val Phe Gln
 130 135 140
 Ala Val Ala Ala Val Leu Ala Leu Leu Val Glu Thr Gly Ala Ile Thr
 145 150 155 160
 Ala Pro Ala Pro Arg Pro Lys Phe Asp Pro Tyr Gly Gln Tyr Gly Arg
 165 170 175
 Tyr Gly Gln Tyr Gly Gln Tyr Gly Val Gln Pro Gly Gly Tyr Tyr Gly
 180 185 190
 Gln Gln Gly Ala Gln Gln Ala Ala Gly Leu Gln Ser Pro Gly Pro Gln
 195 200 205
 Gln Ser Pro Gln Pro Pro Gly Tyr Gly Ser Gln Tyr Gly Gly Tyr Ser
 210 215 220
 Ser Ser Pro Ser Gln Ser Gly Ser Gly Tyr Thr Ala Gln Pro Pro Ala
 225 230 235 240
 Gln Pro Pro Ala Gln Ser Gly Ser Gln Gln Ser His Gln Gly Pro Ser
 245 250 255
 Thr Pro Pro Thr Gly Phe Pro Ser Phe Ser Pro Pro Pro Val Ser
 260 265 270
 Ala Gly Thr Gly Ser Gln Ala Gly Ser Ala Pro Val Asn Tyr Ser Asn
 275 280 285
 Pro Ser Gly Gly Glu Gln Ser Ser Ser Pro Gly Gly Ala Pro Val
 290 295 300

2 INFORMATION FOR SEQ ID NO:93:

- (A) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 39 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

31 SEQUENCE DESCRIPTION (SEQ ID NO:93)

Gly Cys Gly Gln Thr Asp Ala Ala Thr Leu Ala Gln Glu Ala Gly Asn
 1 5 10 15
 Phe Glu Ala Thr Ser Gly Asp Leu Lys Thr Gln Ile
 20 25

32 INFORMATION FOR SEQ ID NO:94

- (A) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 16 amino acids
 (B) TYPE: amino acid

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:94:

Asp Gln Val Glu Ser Thr Ala Gly Ser Leu Gln Gly Gln Trp Arg Gly
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:95:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

Gly Cys Gly Ser Thr Ala Gly Ser Leu Gln Gly Gln Trp Arg Gly Ala
 1 5 10 15
 Ala Gly Thr Ala Ala Gln Ala Ala Val Val Arg
 20 25

(2) INFORMATION FOR SEQ ID NO:96:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

Gly Cys Gly Gly Thr Ala Ala Gln Ala Ala Val Val Arg Phe Gln Glu
 1 5 10 15
 Ala Ala Asn Lys Gln Lys Gln Gln Leu Asp Gln
 20 25

(2) INFORMATION FOR SEQ ID NO:97:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

Gly Cys Gly Ala Asn Lys Gln Lys Gln Gln Leu Asp Gln Ile Ser Thr
 1 5 10 15
 Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg
 20 25

- (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

Gly Cys Gly Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg Ala Asp Glu
 1 5 10 15
 Glu Gln Gln Gln Ala Leu Ser Ser Gln Met Gly Phe
 20 25

(2) INFORMATION FOR SEQ ID NO:99:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 507 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

ATGAAGATGG TGAAATCGAT CGCCGCAGGT CTGACCGCCG CGGCTGCAAT CGGCGCCGCT 60
 GCGGCCCGGTG TGACTTCGAT CATGGCTGGG GCGCCGGTGG TATACCAGAT GCAGCCCGTC 120
 GTCTTGGGCG CGCCACTGCC GTTGGACCGG GCATCCGCCC CTGACGTCCC GACCGCCGCC 180
 CAGTTGACCA GCTTGCTCAA CAGCCTCCCG GATCCCAACG TGTCGTTTGC GAACAAGGGC 240
 AGTCTGTGTC AGGGCGGGCAT CGGGGGCACC GAGGCGCGCA TCGCCGACCA CAAGCTGAAG 300
 AAGGCGCGCG AGCACGGGGA TCTGCCGCTG TCGTTCAGCG TGACGAACAT CCAGCCGGCG 360
 GCGGCGGGTT CGGCCACCGC CGACSTTTCG GTCTGGGGTC CGAAGCTCTC GTCGCCGGTC 420
 ACGCAGAACG TCACSTTCGT GAATCAAGGC AGCTGGATGC TGTACCGCGC ATCGCGGATG 480
 GAGTTGCTGC AGGCGCGAGG GAACTGA 507

(2) INFORMATION FOR SEQ ID NO:100:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 158 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

Met Lys Met Val Lys Ser Ile Ala Ala Gly Leu Thr Ala Ala Ala Ala
 1 5 10 15
 Ile Gly Ala Ala Ala Ala Gly Val Thr Ser Ile Met Ala Gly Gly Pro
 20 25 30
 Val Val Tyr Glu Met Glu Pro Val Val Phe Gly Ala Pro Leu Pro Leu
 35 40 45
 Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Glu Leu Thr Ser
 50 55 60
 Leu Leu Asp Thr Leu Ala Asp Pro Ala Thr Thr Thr Thr Thr Thr Thr

```

          100              105              110
Ser Val Thr Asn Ile Gln Pro Ala Ala Ala Gly Ser Ala Thr Ala Asp
          115              120              125
Val Ser Val Ser Gly Pro Lys Leu Ser Ser Pro Val Thr Gln Asn Val
          130              135              140
Thr Phe Val Asn Gln Gly Gly Trp Met Leu Ser Arg Ala Ser Ala Met
          145              150              155              160
Glu Leu Leu Gln Ala Ala Gly Asn
          165

```

(2) INFORMATION FOR SEQ ID NO:101:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 500 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:101:

```

CGTGGCAATG TCGTTGACCG TCGGGGCCCG GGTGGCCTCC GCAGATCCCG TGGACGCGGT      60
CATTAACACC ACCTGCAATT ACGGGCAGGT AGTAGCTGCG CTCAACGCGA CGGATCCGGG      120
GGCTGCGGCA CAGTTCAACG CCTCACCAGT GCGGCAGTCC TATTGCGCA ATTTCTCTCG      180
CGCACCAGCA CCTCAGCGCG CTGCCATGGC CGCGCAATTG CAAGCTGTGC CGGGGGCGGC      240
ACACTACATC GGCCTTGTCT AGTCGGTTGC CGGCTCCTGC AACAACTATT AAGCCCCATGC      300
GGGCCCCATC CCGCGACCCG GCATCGTCGC CGGGGCTAGG CCAGATTGCC CCGCTCCTCA      360
ACGGGGCGCA TCCCGCGACC CCGCATCGTC GCGGGGGCTA GGCCAGATTG CCGCGCTCCT      420
CAACGGGGCG CATCTCGTGC CGAATTCCTC CAGCCCCGGG GATCCACTAG TTCTAGAGCG      480
GGCGCCACCG CGGTGAGCT

```

(3) INFORMATION FOR SEQ ID NO:102:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 96 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:102:

```

Val Ala Met Ser Leu Thr Val Gly Ala Gly Val Ala Ser Ala Asp Pro
1          5          10          15
Val Asp Ala Val Ile Asn Thr Thr Lys Asn Tyr Gly Gln Val Val Ala
20          25          30
Ala Leu Asn Ala Thr Asp Pro Gly Ala Ala Ala Gln Phe Asn Ala Ser
35          40          45
Pro Val Ala Gln Ser Tyr Leu Arg Asn Phe Leu Ala Ala Pro Pro Pro
50          55          60
Gln Arg Ala Ala Met Ala Ala Gln Leu Gln Ala Val Pro Gly Ala Ala
65          70          75          80
Gln Thr

```

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 154 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:103:

```

ATGACAGAGC AGCACTGGAA TTTCGCGGGT ATCGAGGCCC CGGCAAGCCG AATCCAGGGA      60
AATGTCACGT CCATTCATTC CCTCCTTGAC GAGGGGAAGC AGTCCCTGAC CAAGCTCGCA      120
GCGGCCTGGG GCGGTAGCGG TTCGGAAGCG TACC                                154
  
```

(2) INFORMATION FOR SEQ ID NO:104:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 51 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:104:

```

Met Thr Glu Gln Gln Trp Asn Phe Ala Gly Ile Glu Ala Ala Ala Ser
 1             5             10             15
Ala Ile Gln Gly Asn Val Thr Ser Ile His Ser Leu Leu Asp Glu Gly
      20             25             30
Lys Gln Ser Leu Thr Lys Leu Ala Ala Ala Trp Gly Gly Ser Gly Ser
      35             40             45
Glu Ala Tyr
      50
  
```

(2) INFORMATION FOR SEQ ID NO:105:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 282 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:105:

```

GCTTCTGCGCA TTTCAGGCTG ACTATGAAAG TCGGCTTCGAG GGGGAGGAT TCGTGAAGCT      60
CGAATGCTGGG CGATTAAGCTG AGGTGCATCA TTAGCGGACT TTTCGAGAAG ATCCTGACCG      120
GCTGGAAGAGG CGGACAGCGG GAGGCTGCGT TCGGCTAGG GCTGCTGCGA AATCCCTCGA      180
TAAATTTGCGT CGGGGGCGCG TACAAGGAAG TCGGCTCTGA ATTGCTGCTG TATCTGCTCG      240
AGTTCTGCTG TCTGAGCGCG GAGGAGCGCG TCGGCTGCTG CG                                282
  
```

(2) INFORMATION FOR SEQ ID NO:106:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 154 base pairs

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:106:

GATCGTACCC GTCCGAGTGC TCGGGCCGTT TGAGSATGGA GTGCACGTGT CTTTCGTGAT	60
GGCATACCCA GAGATGTTGG CGGCGGCGGC TGACACCCTG CAGAGCATCG GTGCTACCAC	120
TGTGGCTAGC AATGCCGCTG CGGCGGCCCC GACGACTGGG GTGGTCCCCC CCGCTGCCGA	180
TGAGGTGTCT GCGCTGACTG CGGCGCACTT CGCCGCACAT GCGGCGATGT ATCAGTCCGT	240
GAGCGCTCGG GCTGCTGCGA TTCATGACCA GTTCGTGGCC ACCCTTGCCA GCAGCGCCAG	300
CTCGTATGCG GCCACTGAAG TCGCCAATGC GCGGGCGGTC AGCTAAGCCA GGAACAGTGC	360
GCACGAGAAA GCACGAGAAA TAGGGACACG TAATGGTGGG TTTCGGGGCG TTACCACCGG	420
AGATCAACTC GCGGAGGAGG TACGCGCGGC CGGGTTGCGC CTCGCTGGTG CCGCGCGCTC	480
AGATGTGGGA CAGCGTGGCG AGTGACCTGT TTTCGGCGCG GTCGGCGTTT CAGTGGGTGG	540
TCTGGGTCTT GACGCTGGGG TCGTGGATAG GTTCGTGGCG GGGTCTGATG GTGGCGCGCG	600
CCTCGCGCTA TGTGGCGTGG ATGAGCGTCA CCGCGGGGCA GCGCGAGCTG ACCGCGCGCC	660
AGCTCGGGGT TGCTCGCGCG GCTTACGAGA CCGCGTATGG CCTGACGGTG CCGCGCGCGG	720
TGATCGCGCA GAACCTGCTT GAAGTGATGA TTCTGATAGC GACCAACCTC TTGGGGGCAA	780
AAACCGCGCG CATCGCGCTC AACGAGGCGG AATACGGCGA GATGTGGGCG CAAGACGCGG	840
CGCGCATGTT TCGGTACGCG CCGCGGACCG GCACGCGGAG GCGGACGTTG CTGCGCTTGG	900
AGGAGGCGCG CGACATGACC AGCGCGGCTT GCTTCTGCGA GAGCGCGCGG CCGGTGAGCG	960
AGGCTTGGGA GACCGCGCGG CGGAACCACT TATTAAGCAA TGTGCGCGAG CCGCTGCAAC	1020
AGCTGGCGCA GCGCACGCGG GGCACGACCG GTTCTTCCAA GCTCGGTGGC CTGTGGAAGA	1080
CGGTCTGGCG GCATCGCTCG TCGATGAGCA AGATCTGCTT GATGCGCAAG AACCATGTGT	1140
GAATGATGAA TTGGGTGCTG TCGATGAGCA AGATCTGCTT GATGCGCAAG AACCATGTGT	1200
TTGCGCGCTG TCGGTGCTG TCGATGAGCA AGATCTGCTT GATGCGCAAG AACCATGTGT	1260
CGGTCTGGCG GCATCGCTCG TCGATGAGCA AGATCTGCTT GATGCGCAAG AACCATGTGT	1320
CGGTCTGGCG GCATCGCTCG TCGATGAGCA AGATCTGCTT GATGCGCAAG AACCATGTGT	1380
CGGTCTGGCG GCATCGCTCG TCGATGAGCA AGATCTGCTT GATGCGCAAG AACCATGTGT	1440

TTCCGCGGCC GGCTATGACA ACAGTCAATG TGCATGACAA GTTACAGGTA TTAGGTCCAG 1680
 GTTCAACAAG GAGACAGGCA ACATGGCCTC ACGTTTTATG ACGGATCCGC ACGCGATGCG 1740
 GGACATGGCG GGCCGTTTTG AGGTGCACGC CCAGACGGTG GAGGACGAGG CTCGCCGGAT 1800
 GTGGGCGTCC GCGCAAAACA TTTCCGGTGC GGGCTGGAGT GGCATGGCCG AGGCGACCTC 1860
 GCTAGACACC ATGGCCCAGA TGAATCAGGC GTTTCGCAAC ATCGTGAACA TGCTGCACGG 1920
 GGTGCGTGAC GGGCTGGTTC GCGACSCCAA CAACTACGAG CAGCAAGAGC AGGCCTCCCA 1980
 GCAGATCCTC AGCAGCTAAC GTCAGCCCTC GCAGCACAAT ACTTTTACAA GCGAAGGAGA 2040
 ACAGGTTTGA TGACCATCAA CTATCAATTC GGGGATGTCT ACGCTCACGG CGCCATGATC 2100
 GCGGTCAGG CCGGGTTGCT GGAGGCGGAG CATCAGGCCA TCATTGCTGA TGTGTTGACC 2160
 GCGAGTGAAT TTTGGGGCGG CGCCGGTTCT GCGGCTGCC AGGGGTTTAT TACCCAGTTG 2220
 GGGCTTAAT TCCAGGTGAT CTACGAGCAG GCCAACGCCC ACGGGCAGAA GGTGCAGGCT 2280
 GCGGGCAACA ACATGGCGCA AACCGACAGC GCGCTCGGCT CCAGCTGGGC CTGACACCAG 2340
 GCGAAGGCGA GGGACGTGCT GTACGAGTGA AGTTCCTGCG GTGATCCTTC GGGTGGCAGT 2400
 CTAAGTGGTC AGTGTGCGG TGTGCTGGT TTGCTGCTTG GCGGTTCTT CCGTCTGGT 2460
 CAGTCTCTCT GGGCTCGGG TGAGGACCTC GAGGCCCAGG TAGCCCGCTC CTTCGATCCA 2520
 TTCTCTCTCT TGTTCGGCGA GGACGCTCC GACGAGGCGG ATGATCGAGG CCGGTCGCG 2580
 GAAGATGCCC AGGACGTGG TTCCGCTCG TACCTCTCG TTGAGCGCTT CCGGGGCTT 2640
 GTTGGACCAG ATTTGGCGCT AGATCTCTT GCGGAAGGCG GTGAACGCCA GCAGGTCTCT 2700
 GCGGCGCGTG TCGAGGTGCT GGGCCACCGC GGGGAGTTTG TCGCTCAGAG CGTCGAGTAC 2760
 TGATGATAT TGGCAACAA GTGATTCGG GTGGGCTG TGTAGATG AGTGCAGCA 2820
 GTTGGCAGC GAGGCGCAG AGGTTTCTG GTTCTCTG ATAGATTG TTGCTAGT 2880
 GTTGTGCGAG CGCTGCGAG GCGTGTGCG GAGGTGCG TGTATCGCG CCAGCAGGCC 2940
 GTTGTGCGAG TCGCTGGTGA GCGGCGCAG GCGGACAGG TCGGCGCGA CCAGGTGCG 3000
 GAAGAAGGCC AGCCAGCCGG TCGCTCTCT GCGGAGGTG AGCTGGATG CCAGGATC 3054

INFORMATION FOR SEQ. 1: NUC. SEQ.

123

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:107:

Met Val Asp Phe Gly Ala Leu Pro Pro Glu Ile Asn Ser Ala Arg Met
 1 5 10 15
 Tyr Ala Gly Pro Gly Ser Ala Ser Leu Val Ala Ala Ala Gln Met Trp
 20 25 30
 Asp Ser Val Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Gln Ser
 35 40 45
 Val Val Trp Gly Leu Thr Val Gly Ser Trp Ile Gly Ser Ser Ala Gly
 50 55 60
 Leu Met Val Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr
 65 70 75 80
 Ala Gly Gln Ala Glu Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala
 85 90 95
 Ala Tyr Glu Thr Ala Tyr Gly Leu Thr Val Pro Pro Pro Val Ile Ala
 100 105 110
 Glu Asn Arg Ala Glu Leu Met Ile Leu Ile Ala Thr Asn Leu Leu Gly
 115 120 125
 Gln Asn Thr Pro Ala Ile Ala Val Asn Glu Ala Glu Tyr Gly Glu Met
 130 135 140
 Trp Ala Gln Asp Ala Ala Ala Met Phe Gly Tyr Ala Ala Ala Thr Ala
 145 150 155 160
 Thr Ala Thr Ala Thr Leu Leu Pro Phe Glu Glu Ala Pro Glu Met Thr
 165 170 175
 Ser Ala Gly Gly Leu Leu Glu Gln Ala Ala Ala Val Glu Glu Ala Ser
 180 185 190
 Asp Thr Ala Ala Ala Asn Ala Leu Met Asn Asn Val Pro Gln Ala Leu
 195 200 205
 Gln Gln Leu Ala Ala Pro Thr Gln Gly Thr Thr Pro Ser Ser Lys Leu
 210 215 220
 Gly Gly Leu Trp Lys Thr Val Ser Pro His Arg Ser Pro Ile Ser Asn
 225 230 235 240
 Met Val Ser Met Ala Asn Asn His Met Ser Met Thr Asn Ser Gly Val
 245

124

Ala Ala Ala Gln Ala Val Gln Thr Ala Ala Gln Asn Gly Val Arg Ala
 275 280 285

Met Ser Ser Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu Gly Gly Gly
 290 295 300

Val Ala Ala Asn Leu Gly Arg Ala Ala Ser Val Gly Ser Leu Ser Val
 305 310 315 320

Pro Gln Ala Trp Ala Ala Ala Asn Gln Ala Val Thr Pro Ala Ala Arg
 325 330 335

Ala Leu Pro Leu Thr Ser Leu Thr Ser Ala Ala Glu Arg Gly Pro Gly
 340 345 350

Gln Met Leu Gly Gly Leu Pro Val Gly Gln Met Gly Ala Arg Ala Gly
 355 360 365

Gly Gly Leu Ser Gly Val Leu Arg Val Pro Pro Arg Pro Tyr Val Met
 370 375 380

Pro His Ser Pro Ala Ala Gly
 385 390

(2) INFORMATION FOR SEQ ID NO:108:

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1725 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:108:

TACGTCAGCA CCGGCGGTGC AGGGCTGGAG GTGCTGCTT TTTGATCTGC GGTCAAGGTG 60

AGGTGCGCTG GCGTGTGGCT GGCCTGGATG CAGATTCGAT CCGGCTCTTT AGTGCAACTA 120

ATTTCCTTGA AGTGGCTGCG AGGTATAGGA GTTGAATAT GTTAAATSTA GCGTTCACGC 180

TTTGTGGCGG TCGATTTCGG TGGAGCACTG ATCAATAAAT TTGGCGCTGC GCGCCAGGCG 240

TTGATGAGA TCGCTTGANT AATATCAAT TTGAATATC TGGGCGGATG CTGGGGGTAA 300

ATGAGGAGGA GCAGCGCTGT GTTGTACTGT TGAATGAG AGTTTGGCGG CCGCGGCTGG 360

TAAATTCCTT TCGCTGGGGG CAAGCTGAAA GGCTAGCAAT GCGGCGGCGG CCGTGGCGAC 420

TAATAGCTG TTGCGCGCGG TGAAGAGCA GTTGTGCTT TTCTTGGCA CACAATTCGG 480

TAATATATC TGAAGCTATC AATATATC

CCGGGCGGTTA CCACCGGAGA TCAACTCCGC GAGGATGTAC GCCGGCCCCG GTTCGGCCTC 720
 GCTGGTGGCC GCCCGAAGA TGTGGGACAG CGTGGCGAGT GACCTGTTTT CGGCCGCGTC 780
 GCGGTTTCAG TCGGTGGTCT GGGGTCTGAC GGTGGGGTCC TGGATAGGTT CGTCGGCGGG 840
 TCTGATGGCG GCGGCGGCCT CGCGTATGT GCGGTGGATG AGCGTCACCG CGGGGCAGGC 900
 CCAGCTGACC GCCGCCGAGG TCGGGTTTC TCGGGCGGCC TACGAGACAG CGTATAGGCT 960
 GACGGTGGCC CGCCCGGTGA TCGCCGAGAA CCGTACCGAA CTGATGACCG TGACCGCGAC 1020
 CAACCTCTTG GGGCAAAACA CGCCGGCGAT CGAGGCCAAT CAGGCCGCAT ACAGCCAGAT 1080
 GTGGGGCCAA GACCGGAGG CGATGTATGG CTACCGCGCC ACGGCGGCGA CGGCGACCGA 1140
 GCGCTTGCTG CCGTTCCAGG AGCGCCCACT GATCAGCAAC CGCGGGCGGC TCCTTGAGCA 1200
 GCGCGTCCG GTGAGGAGG CCATCGACAC CGCCCGCGCG AACCAAGTTGA TGAACAATGT 1260
 GCGCCAAAGG CTGCAACAGC TGGCCAGCC AGCGCAGGGC GTCGTACCTT CTTCGAAGCT 1320
 GGGTGGGCTG TGGACGGCG TCTCGCGCA TGTGTGCGG CTCAGCAACG TCAGTTGGAT 1380
 AGCCAAACAC CACATGTGCA TGATGGGCAC GGGTGTGTG ATGACCAACA CATTGCACTC 1440
 GATGTTGAAG GCGTTAGCTG CGCGGGCGGC TCAGGCGCTG GAAACCGCGG CGGAAAACGG 1500
 GGTGTGGCG ATGAGCTCG TGGCCAGCCA GCTGGGTTCC TCGCTGGGTT CTTCGGGTCT 1560
 GCGCGCTCG GTGGCGGCA ATTGGGTGCG GCGGTGCTG CTCGGTTGCT TGTGGTGGC 1620
 TCGAGCTCG CGCGCGGCA AGCAGCGGCT CAGCGGCGG CGCGCGGCG TCGCGGTGAC 1680
 TAGGCTGACC AGCGCGGCG AAACCGCGCG CGGACACATG GTGGC 1725

2. INFORMATION FOR SEQ ID NO:109.

1. SEQUENCE CHARACTERISTICS

- A. LENGTH: 194 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS:
- D. TOPOLOGY: linear

3. SEQUENCE DESCRIPTION: SEQ ID NO:109

Val Val Asp Phe Gly Ala Leu Pro Pro Gly Ile Asn Ser Ala Arg Met
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

126

Val Val Trp Gly Leu Thr Val Gly Ser Trp Ile Gly Ser Ser Ala Gly
 50 55 60
 Leu Met Ala Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr
 65 70 75 80
 Ala Gly Gln Ala Gln Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala
 85 90 95
 Ala Tyr Glu Thr Ala Tyr Arg Leu Thr Val Pro Pro Pro Val Ile Ala
 100 105 110
 Glu Asn Arg Thr Glu Leu Met Thr Leu Thr Ala Thr Asn Leu Leu Gly
 115 120 125
 Gln Asn Thr Pro Ala Ile Glu Ala Asn Gln Ala Ala Tyr Ser Gln Met
 130 135 140
 Trp Gly Gln Asp Ala Glu Ala Met Tyr Gly Tyr Ala Ala Thr Ala Ala
 145 150 155 160
 Thr Ala Thr Glu Ala Leu Leu Pro Phe Glu Asp Ala Pro Leu Ile Thr
 165 170 175
 Asn Pro Gly Gly Leu Leu Glu Gln Ala Val Ala Val Glu Glu Ala Ile
 180 185 190
 Asp Thr Ala Ala Ala Asn Gln Leu Met Asn Asn Val Pro Gln Ala Leu
 195 200 205
 Gln Gln Leu Ala Gln Pro Ala Gln Gly Val Val Pro Ser Ser Lys Leu
 210 215 220
 Gly Gly Leu Trp Thr Ala Val Ser Pro His Leu Ser Pro Leu Ser Asn
 225 230 235 240
 Val Ser Ser Ile Ala Asn Asn His Met Ser Met Met Gly Thr Gly Val
 245 250 255
 Ser Met Thr Asn Thr Leu His Ser Met Leu Lys Gly Leu Ala Pro Ala
 260 265 270
 Ala Ala Gln Ala Val Gln Thr Ala Ala Glu Asn Gly Val Trp Ala Met
 275 280 285
 Ser Ser Leu Gly Ser Gln Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu
 290 295 300
 Gly Ala Gly Val Ala Ala Asn Leu Gln Val Ala Ala Ser Val Gly
 305

Ala Ala Arg Ala Leu Pro Leu Thr Ser Leu Thr Ser Ala Ala Gln Thr
 340 345 350

Ala Pro Gly His Met Leu Gly
 355

(2) INFORMATION FOR SEQ ID NO:110:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3027 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:110:

AGTTCACTCG AGAATGATAC TGACGGGGCTG TATCCAGCAT GGCTGAGACA ACCGAACCAAT 50
 GGTGGGACCG GGGGACATCG CAAGCCGACG CGATGGGCTT GGCCGCCGAA GCCGAAGCCG 120
 GCGAAGCCGA AGCCCTGGGC GCGCGGGCGG GCGCCCGTGT CCGTGCCGCC CCGTTGAAGC 180
 GTGAGGGCGT GCGCATGGCC CGAGCCGAGG ACGAGAACST CCCCAGGAT ATGCAGACTG 240
 GGAAGACCGC GAAGACTATG ACGACTATGA CGACTATGAG GCCGCAGACC AGGAGGCCGC 300
 AGGTTGGGGA TCGTGGCGAC GCGGCTTGGG GGTGCGGTTA CCAAGACTGT CCACGATTGC 360
 CATGCGGGCC GGAGTCTCA TCATCTGCGG TTTCACCGGG CTCAGCGGAT ACATTGTGTG 420
 CCAACACCAT GAGGCCACCG AACCTCAGCA GCGCGCGCGG GCGTTGCGCG CCGGAGCCAA 480
 CCAAGTGTTC ATCAACATCA CTCCTCTGGA ATTGACGAGG GCCAAGAAG AGCTGGGCGC 540
 TGTGATCGAC AGCTGACCG GCGAATTGAG CGATGACTTC CAGCAGCGGG CAGCTGATTT 600
 CAGCAGGTT GTGGAACAGT CCAAGTGGT TACCGAAGCG ACGGTGAACG CGACAGCCGT 660
 CCAATGATG AATGAGGATT CGCTCTGCTT TTTCTGTTG GCGACTTCAG GGGTACGAA 720
 TTTTCTGTTT TCGAAGAAAT AATGATCTG ATCTCTGTT TAACTGACCG TGACTTAAGA 780
 GCGGCGACAG TAAAGATTT TGAATTTCA TTCTGTA TCGACGATGA CTTAGCGGAT 840
 GTCAAGACCG AAACCACTGA GCGGAGCGAA TCTCTTAAA TCGACTCAGC GCGAGCGGAA 900
 GCGCTGATTT GCGGAGCTGA GCGATTTGAG ACGAATTTT TACCGGAATC TACCGCGCAG 960
 AAATCTAAAT GCGAATTTA TCTCTCTTA ATCTATTTA ATTGAAAGT GGTTCGCTT 1020

AGGTCGCACC TCGCCGGCGA TTCTCTGTCC TATACGACCA GTTCACGCAG CAGATCGTGG 1260
CTCCGGCGGC CAAACAGAAG TCACTGAAAA CCACCGCCAA GGTGGTGCGC GCGGCCGTGT 1320
CGGAGCTACA TCCGGATTCC GCCGTCGTTC TGGTTTTTGT CGACCAGAGC ACTACCAGTA 1380
AGGACAGCCC CAATCCGTCG ATGGCGGCCA GCAGCGTGAT GGTGACCCTA GCCAAGGTCC 1440
ACGGCAATTG GCTGATCAUC AAGTTACCCC CGGTTTAGGT TGCCGTAGGC GGTGCGCAAG 1500
TCTGACGGGG GCGCGGGTGG CTGCTCGTGC GAGATACCGG CCGTTCTCCG GACAATCAGC 1560
GCCCGACCTC AAACAGATCT CGGCCGCTGT CTAATCGGCC GGGTTATTTA AGATTAGTTG 1620
CCACTGTATT TACCTGATCT TCAGATTGTT CAGCTGGATT TAGCTTCCCG GCAGGGCGCG 1680
TGTGCACTT TGCATCTGGG GTTGAGACTA CTTGAGAGAA TTGAGCTGT TCGCGACGTT 1740
GTTTCTGTTC CATCATTTGG GCTAGTTATG GCCGAGCGGA AGGATTATCG AAGTGGTGG 1800
CTTCGGGGCG TTACCACCGG AGATCAACTC CGCGAGGATG TACGCCGGCC CGGTTTCGGC 1860
CTCGCTGGTG GCCCGCGCGA AGATGTGGGA CAGCCTGGCG AGTGACCTGT TTTCGGCCGC 1920
GTCCCGCTTT CAGTGGGTGG TCTGGGGTCT GAUGACGGGA TCGTGGATAG GTTCGTCCGC 1980
TGGTGTGATG GTGGCGGCGG CCTCGCGCTA TGTGGCGTGC ATGAGCGTCA CCGCGGGGCA 2040
TCCCGACCTC ACCCGCGGCC AGGTCCGGGT TGGTGGCGG GCTTACGAGA CGGCSTATGG 2100
TCTGACGGTG CCCCCCGCGG TGATCCCGGA GAAGCTGCT GAAGTGTGTA TTCTGATAG 2160
TATCAACCTC TTGGGGGAAA AACCGCGCGC GATGGCGGT AACGAGCGCG AATACGGGGA 2220
TATGTGGGCG CAAGACGCGG CCGCATCTT TGGCTAGCGG GCGACGGCGG CGACGGCGAC 2280
TACGGCGTTG CTCCGTTTC AGGACCGCGC ACTGACGAG AACCGCGCGG GGTCTCTTGA 2340
TCACTCTCTG CCGCTGAGG AAGGATGGA AGGCTGCGG TCAACGCTT TGATGAACAA 2400
TCTTCTGAAA TGGTGGAA ATATAGTAA TGGGAGAAA AGGATCTT TCTTGGACCA 2460
ATTAATGAA TCTGGAAG CCACTGCGCC GATCTGTCT TGGTGGCA ACATCTGTCT 2520
TATCTGAA AACCACTCT GATGAGAA CTGGGTCTT TGATGGCA CCACCTTGA 2580
CTTAATCTT AAGGCTTCT CCGCTGCGG TGGTGGCG GTGAAAGCG CCGCGGAAA 2640
TGGTGGCA TATTAATCT TCTGCGAA TCACTGCTT TGGTGGCG GTTCTTGGCA

GACCAGCCTG ACCAGCGCCG CCCAAACCGC CCGCGGACAC ATGCTGGGCG GGCTACCGCT 2880
 GGGGCAACTG ACCAATAGCG GCGGCGGGTT CGGCGGGGTT AGCAATGCGT TCGCGATGCC 2940
 GCGCGGGGCG TACGTAATGC CCGTGTGCC CGCGCGCGG TAACGCCGAT CCGCACGCA 3000
 TCGGGGCCCT CTATGCGGCC AGCGATC 3027

(2) INFORMATION FOR SEQ ID NO:111:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 396 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:111:

Val Val Asp Phe Gly Ala Leu Pro Pro Glu Ile Asn Ser Ala Arg Met
 1 5 10 15
 Tyr Ala Gly Pro Gly Ser Ala Ser Leu Val Ala Ala Lys Met Trp
 20 25 30
 Asp Ser Val Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Gln Ser
 35 40 45
 Val Val Trp Gly Leu Thr Thr Gly Ser Trp Ile Gly Ser Ser Ala Gly
 50 55 60
 Leu Met Val Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr
 65 70 75 80
 Ala Gly Gln Ala Gln Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala
 85 90 95
 Ala Tyr Gln Thr Ala Tyr Gly Leu Thr Val Pro Pro Pro Val Ile Ala
 100 105 110
 His Asn Arg Ala Ile Leu Met Ile Leu Ile Ala Thr Asn Leu Leu Gln
 115 120 125
 Gln Asn Thr Pro Ala Ile Ala Val Asn Ala Ala Gln Tyr Gly Gln Met
 130 135 140
 Trp Ala Gln Asp Ala Ala Ala Met Phe Gly Tyr Ala Ala Thr Ala Ala
 145 150 155 160
 Thr Ala Thr Gln Val Leu Leu Pro Pro Ala Asp Ala Pro Leu Thr Thr

130

195 200 205

Gln Gln Leu Ala Gln Pro Thr Lys Ser Ile Trp Pro Phe Asp Gln Leu
210 215 220

Ser Glu Leu Trp Lys Ala Ile Ser Pro His Leu Ser Pro Leu Ser Asn
225 230 235 240

Ile Val Ser Met Leu Asn Asn His Val Ser Met Thr Asn Ser Gly Val
245 250 255

Ser Met Ala Ser Thr Leu His Ser Met Leu Lys Gly Phe Ala Pro Ala
260 265 270

Ala Ala Gln Ala Val Glu Thr Ala Ala Gln Asn Gly Val Gln Ala Met
275 280 285

Ser Ser Leu Gly Ser Gln Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu
290 295 300

Gly Ala Gly Val Ala Ala Asn Leu Gly Arg Ala Ala Ser Val Gly Ser
305 310 315 320

Leu Ser Val Pro Gln Ala Trp Ala Ala Ala Asn Gln Ala Val Thr Pro
325 330 335

Ala Ala Arg Ala Leu Pro Leu Thr Ser Leu Thr Ser Ala Ala Gln Thr
340 345 350

Ala Pro Gly His Met Leu Gly Gly Leu Pro Leu Gly Gln Leu Thr Asn
355 360 365

Ser Gly Gly Gly Phe Gly Gly Val Ser Asn Ala Leu Arg Met Pro Pro
370 375 380

Arg Ala Tyr Val Met Pro Arg Val Pro Ala Ala Gly
385 390 395

INFORMATION FOR SEC ID NO 111

1. FREQUENCY CHARACTERISTICS:

LENGTH 101 mm

B TYPE: nucleic acid

STRANDEDNESS. Jangle

7. TOPOLOGY Linear:

X1 SEQUENCE DESCRIPTION SEQ TO NO 112

INTERNATIONAL AGREEMENTS AT WASHINGTON, APRIL 22, 1958, CONCERNING AIRCRAFT

GGTACAAACC GCGTCAACAC AGGCCAAGAC CCGTGCGATG CAGGCGACGG CGCAAGCCGC 300
 GGCATACACC CAGGCCATGG CCACGACGCC GTGCGTGCCG GAGATCGCCG CCAACCACAT 360
 CACCCAGGCC GTCCTTACGG CCACCAACTT CTTGGGTATC AACACGATCC CGATCGCGTT 420
 GACCGAGATG GATTATTTC A TCGTATGTG GAACCAGGCA GCCCTGGCAA TGGAGGTCTA 480
 CCAGGCCGAG ACCGCGGTTA ACACGCTTTT CGAGAAGCTC GAGCCGATGG CGTCGATCCT 540
 TGATCCCGGC GCGAGCCAGA GCACGACGAA CCGGATCTTC GGAATGCCCT CCCCTGGCAG 600
 CTCAACACCG GTTGGCCAGT TGCCGCGGGC GGCTACCCAG ACCCTCGGCC AACTGGGTGA 660
 GATGAGCGGC CGGATGCAGC AGCTGACCCA GCGGCTGCLC CAGGTGACGT CGTTGTTTCA 720
 CCAGGTGGGC GGCACCGGCC GCGGCAACCC AGCCGACGAG GAAGCGGCCG AGATGGGCGT 780
 GCTCGGCACC AGTCGGTGT CGAACCATCC GCTCGGTGGT GGATCAGGCC CCAGCGCGGG 840
 CCGCGGCGTG CTGCGCGCGG AGTCGGTACC TGGCGCAGGT GGGTCGTTGA CCGGCACGCC 900
 GCTGATGTCT CAGGTGATCG AAAAGCCGGT TGCCCCCTCG GTGATGCGCG CGGCTGCTGC 960
 CCGATCGTCT GCGACCGGTG GCGCGGCTCG GGTGGGTGCG GGAGCGATGG GCGAGGCTGC 1020
 GGAATCGCGC GGCTGACCA GCGCGGCTCT GGTGCGGCGC GCACCGCTCG CGCAGGAGCG 1080
 TGAAGAAGAC GACGAGGAGC ACTGGGACGA AGAGGAGGAC TGGTGAGCTC CGGTAAATGAC 1140
 AACGAACTTC CCGGCCACCC GGGCGCGAAG ACTTCCCAAC ATTTTGGCGA GGAAGGTAAA 1200
 GAGAGAAAGT AGTCAGCAT GCGAGAGATG AAGAGCGATG CCGCTACCTT CCGGCAGGAG 1260
 CGAGTAATT TCGAGCGGAT CTGCGGCGAC CTGAAAACCC AGATCGACCA GGTGGAGTCC 1320
 ACCGTAGGTT CATTGACGGT CAGTGGGCGC GCGCGCGCGG GGACCGCGCG CCAGGCGTCC 1380
 TTGTTGCTCT TCGAAGAAAT ATTAATAAAG GAGAAGCAGT AACTCGACCA GATCTCGACG 1440
 AATATGCTCT AGCGTGGCTT TGAATACTCT AGGCTGAAAT AGGAGGAGCA CCAAGCGCTG 1500
 TCGTGGGAAA TCGGCTTCTG AGCTGTAAAT AGGAAAAGAA AGGAGGCAAA AACATGACAG 1560
 AGCAGCACTG GAATTTGCGT GGTATCGAGG CCGCGGCAAG TCGAATGCAAG CGAAAT 1616

2. INFORMATION FOR SEQ ID NO 113:

SEQUENCE CHARACTERISTICS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

CTACTGGATG GGACCATGGC CATTCTCTGC AGTCTCACTG CTTTCTGTGT TGACATTTTG 60
 GCACGCCGGC GGAAACGAAG CACTGGGGTC GAAGAACGGC TCGGCTGCCA TATCGTCCCG 120
 AGCTTCCATA CCTTCGTGCG GCCGGAAGAG CTTGTCTAG TCGGCGCCA TGACAACCTC 180
 TCAGAGTGCG CTCAAACGTA TAAACACGAG AAAGGGCGAG ACCGACGGAA GGTGGAACTC 240
 GCGCGATCCC GTGTTTCGCT ATTCTACGCG AACTCGGCGT TGGCCTATGC GAACATCCCA 300
 GTGACGTTGC CTTGGGTGCA AGCCATTGCC TGACCGGCTT CCGTATCGT CCGCGCCAGG 360
 TTCTGCAGCG CGTGTTCAG CTCGGTAGCC GTGGCGTCCC ATTTTGTCTG GACACCCTGG 420
 TACGCTTCG AA 480

D. INFORMATION FOR SEQ ID NO:114:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 368 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

Met Leu Trp His Ala Met Pro Pro Glu Xaa Asn Thr Ala Arg Leu Met 1 5 10 15
 Ala Gly Ala Gly Pro Ala Pro Met Leu Ala Ala Ala Ala Gly Trp Glu 20 25 30
 Thr Leu Ser Ala Ala Leu Asp Ala Gln Ala Val Glu Leu Thr Ala Arg 35 40 45
 Leu Asn Ser Leu Gly Gln Ala Trp Thr Gly Gly Gly Ser Asp Lys Ala 50 55 60
 Leu Ala Ala Ala Thr Pro Met Val Val Thr Leu Gln Thr Ala Ser Thr 65 70 75 80
 His Ala Lys Thr Arg Ala Met Gln Ala Thr Ala Gln Ala Ala Tyr 85 90 95
 Thr Gln Ala Met Ala Thr Thr Pro Ser Leu Pro Gln Ile Ala Ala Asn 100 105 110
 His Ile Thr Gln Ala Val Leu Thr Val Thr Asn Thr Pro 115 120 125

133

Asn Gln Ala Ala Leu Ala Met Glu Val Tyr Gln Ala Glu Thr Ala Val
 145 150 155 160
 Asn Thr Leu Phe Glu Lys Leu Glu Pro Met Ala Ser Ile Leu Asp Pro
 165 170 175
 Gly Ala Ser Gln Ser Thr Thr Asn Pro Ile Phe Gly Met Pro Ser Pro
 180 185 190
 Gly Ser Ser Thr Pro Val Gly Gln Leu Pro Pro Ala Ala Thr Gln Thr
 195 200 205
 Leu Gly Gln Leu Gly Glu Met Ser Gly Pro Met Gln Gln Leu Thr Gln
 210 215 220
 Pro Leu Gln Gln Val Thr Ser Leu Phe Ser Gln Val Gly Gly Thr Gly
 225 230 235 240
 Gly Gly Asn Pro Ala Asp Glu Gln Ala Ala Gln Met Gly Leu Leu Gly
 245 250 255
 Thr Ser Pro Leu Ser Asn His Pro Leu Ala Gly Gly Ser Gly Pro Ser
 260 265 270
 Ala Gly Ala Gly Leu Leu Arg Ala Gln Ser Leu Pro Gly Ala Gly Gly
 275 280 285
 Ser Leu Thr Arg Thr Pro Leu Met Ser Gln Leu Ile Gln Lys Pro Val
 290 295 300
 Ala Pro Ser Val Met Pro Ala Ala Ala Ala Gly Ser Ser Ala Thr Gly
 305 310 315 320
 Gly Ala Ala Pro Val Gly Ala Gly Ala Met Gly Gln Gly Ala Gln Ser
 325 330 335
 Gly Gly Ser Thr Arg Pro Gly Leu Val Ala Pro Ala Pro Leu Ala Gln
 340 345 350
 Glu Asp His Ala Asp Asp His Asp Asp Thr Asp Ile His Asp Asp Trp
 355 360 365 370 375 380 385 390 395

4. INFORMATION FOR SEQ ID NO: 113

1. SEQUENCE CHARACTERISTICS

- A. LENGTH: 100 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS:
- D. TOPOLOGY:

Asn Phe Glu Arg Ile Ser Gly Asp Leu Lys Thr Gln Ile Asp Gln Val
 20 25 30
 Glu Ser Thr Ala Gly Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly
 35 40 45
 Thr Ala Ala Gln Ala Ala Val Val Arg Phe Gln Glu Ala Ala Asn Lys
 50 55 60
 Gln Lys Gln Glu Leu Asp Glu Ile Ser Thr Asn Ile Arg Gln Ala Gly
 65 70 75 80
 Val Gln Tyr Ser Arg Ala Asp Glu Glu Gln Gln Ala Leu Ser Ser
 85 90 95
 Gln Met Gly Phe
 100

(E) INFORMATION FOR SEQ ID NO:116:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 396 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:116:

GATCTCCGGG GACCTGAAAA CCCAGATCGA CCAGGTGGAG TCGACGGCAG GTTCGTTGCA 60
 GGGCCAGTGG GCGGGCCCGG CGGGGACGGG CCCCCAGGGG GCGGTGGTGG GCTTCCAAGA 120
 ACCAGCCCAT AAGCAGAAAG AGGAATTGTA CGAGATCTCG ACCAATATTC CTCAGGCCCG 180
 GTTCCAATAC TCGAGGGCGG ACCAGGAGCA GCAGGAGGGG CTGTCTCTGG AAATGGGCTT 240
 GTGACCCCGT AATACGAAAA GAAACGGAGG AAAACATGA CAGAGCAGCA GTGGAATTTG 300
 CCGGCTATCG AGGCGGCGG AAGCCCAATC AAGCGAATG TCAGCTTCAT TCAATCCGTC 360
 TTGAGGAGG CGAAGCAATC CTGAGTAAAT TTGGA 396

(F) INFORMATION FOR SEQ ID NO:117:

- (1) SEQUENCE CHARACTERISTICS
- A LENGTH: 80 amino acids
 - B TYPE: amino acid
 - C STRANDEDNESS: single
 - D TOPOLOGY: linear

Gly Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln
 20 25 30
 Ala Ala Val Val Arg Phe Gln Glu Ala Ala Asn Lys Gln Lys Gln Glu
 35 40 45
 Leu Asp Glu Ile Ser Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser
 50 55 60
 Arg Ala Asp Glu Glu Gln Gln Gln Ala Leu Ser Ser Gln Met Gly Phe
 65 70 75 80

(2) INFORMATION FOR SEQ ID NO:118:

1. SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 387 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:118:

GTGGATCCCG ATCCCGTGTT TCGCTATTCT ACACGAACCT GGCCTTGCCC TATGCGAACA 60
 TCCGAGTGAC GTTGCCTTCG GTGAAGCCCA TTCCCTGACC GGCCTGCTG ATCCTCCGCG 120
 CCAGGTTCTG CAGCGGCTTG TTCAGCTCGG TAGCCGTGGC GTCCCATTTT TCCTGCACAC 180
 CCTCGTAGCG TTGGAACCG CTACCGCCCG AGCCCGCTGC CAGCTTGCTC AGGGACTGCT 240
 TCCCTCTCTC AAGGAGGGAA TGAATGGAGG TGACATTTCC CTGGATTGCG TTGCGCCCGG 300
 TCTGATAGCG TCGGAATTC CACTGCTGCT CTGCTATGTT TTGCTGCTT TCTTTTTCGT 360
 ATTAGCGGGT CAGAAGCCCA TTGCGA 387

2. INFORMATION FOR SEQ ID NO:119:

1. SEQUENCE CHARACTERISTICS:
 A. LENGTH: 387 base pairs
 B. TYPE: nucleic acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

X1. SEQUENCE DESCRIPTION: SEQ ID NO:119:

TTCGAGAGG ATCTGCTTTT ACCGAGGCTG TTGCTGAGG GATCCTTTCT TATTTGAGT
 TCTGCTGCTG ATCTGCTTTT TCTGCTGCTG TCTGCTGCTG TCTGCTGCTG

GGCGGGGGTT CCGCGATTGG CATCTTGGCC CA

272

(2) INFORMATION FOR SEQ ID NO:120:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

Asp Pro Val Asp Ala Val Ile Asn Thr Thr Cys Asn Tyr Gly Gln Val
 1 5 10 15
 Val Ala Ala Leu
 20

(2) INFORMATION FOR SEQ ID NO:121:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

Ala Val Glu Ser Gly Met Leu Ala Leu Gly Thr Pro Ala Pro Ser
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:122:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 13 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

Ala Ala Met Lys Pro Asp Ile His Asn Gly Pro Leu Glu Ala Ala Lys
 1 5 10 15
 Glu Gly Arg

(2) INFORMATION FOR SEQ ID NO:123:

- (i) SEQUENCE CHARACTERISTICS:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:123:

Tyr Tyr Trp Cys Pro Gly Gln Pro Phe Asp Pro Ala Trp Gly Pro
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:124:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 14 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:124:

Asp Ile Gly Ser Glu Ser Thr Glu Asp Gln Gln Xaa Ala Val
1 5 10

(2) INFORMATION FOR SEQ ID NO:125:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 13 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:125:

Ala Glu Glu Ser Ile Ser Thr Xaa Glu Xaa Ile Val Pro
1 5 10

(2) INFORMATION FOR SEQ ID NO:126:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 17 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:126:

Asp Pro Glu Pro Ala Pro Pro Val Pro Thr Thr Ala Ala Ser Pro Pro
1 5 10 15

Ser

(2) INFORMATION FOR SEQ ID NO:127:

- (1) SEQUENCE CHARACTERISTICS:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

Ala Pro Lys Thr Tyr Xaa Glu Glu Leu Lys Gly Thr Asp Thr Gly
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:128:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Gln Leu Thr Ser
1 5 10 15
Leu Leu Asn Ser Leu Ala Asp Pro Asn Val Ser Phe Ala Asn
20 25 30

(2) INFORMATION FOR SEQ ID NO:129:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

Asp Pro Pro Asp Pro His Gln Xaa Asp Met Thr Lys Gly Tyr Tyr Pro
5 10 15
Gly Gly Arg Arg Xaa Phe
20

(2) INFORMATION FOR SEQ ID NO:130:

(i) SEQUENCE CHARACTERISTICS:

- A LENGTH: 7 amino acids
- B TYPE: amino acid
- C STRANDEDNESS
- D TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

Asp Asp Gly Tyr Thr Ser Gly
1 5

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: /note= "The Second Residue Can Be Either a Pro or Thr"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

Xaa Xaa Gly Phe Thr Gly Pro Gln Phe Tyr
 1 5 10

(2) INFORMATION FOR SEQ ID NO:132:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: /note= "The Third Residue Can Be Either a Gln or Leu"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

Xaa Pro Xaa Val Thr Ala Tyr Ala Gly
 1 5

(2) INFORMATION FOR SEQ ID NO:133:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

Xaa Xaa Val Gln Lys Pro Phe Lys Arg
 1

(2) INFORMATION FOR SEQ ID NO:134:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO:135:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:135:

Ala	Gly	Asp	Thr	Xaa	Ile	Tyr	Ile	Val	Gly	Asn	Leu	Thr	Ala	Asp
1				5					10				15	

(2) INFORMATION FOR SEQ ID NO:136:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:136:

Ala	Pro	Glu	Ser	Gly	Ala	Gly	Leu	Gly	Gly	Thr	Val	Gln	Ala	Gly
1				5					10				15	

(2) INFORMATION FOR SEQ ID NO:137:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 21 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:137:

Xaa	Tyr	Ile	Ala	Tyr	Xaa	Thr	Thr	Ala	Gly	Ile	Val	Pro	Gly	Lys	Ile
1				5					10					15	
Asn	Val	His	Leu	Val											

(2) INFORMATION FOR SEQ ID NO:138:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 882 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO:10:

2. DEFINITION OF CHARACTERISTICS

- A: LENGTH: 815 base pairs
B: TYPE: nucleic acid
C: STRANDEDNESS: single
D: TOPOLOGY: linear

1. *Journal of the American Medical Association*, 1997; 277: 1039-1043.

Table 1. *Salmonella* serotypes and their associated diseases

GGAGTGGCGG GATTAGGGCA CTCACCGGCG GAAGCAGCGG AAGTACTCTG GGTACCCGAG 480
 CACTTCGAGA CGACACUACA GCACGGGGTT GTGGCGGGAC TGGGTGAAAT GGCAGCCGAT 540
 AGCGECTAGG TGTCGGCTGC GGTCAACCTC GATCATGATG TCGAGGTGAC CSTGACCGCG 600
 CCCCCTGAAG GAGGCGCTGA ACTCGGCGTT GAGCGGATCG GCGATCGGTT GGGGCAGTGC 660
 CCAGSCCAAT ACGGGGATAC CGGTGTCTNA AGCGGCGCGG AGCGCAGGTT CGGTTCGCGG 720
 ACGTGTCTCG GGTGCGCTTG TTACGCGCTT GTGTCGAAC ACGAGTAGGA GGTCTGCTCG 780
 GCGGAGGGCA TCGAGGAGG GTTCTGTCAG CTCCT 815

(12) INFORMATION FOR SEQ ID NO:140:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1152 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

ATGAGTGGCG GGTGAGCTC TCAGATCAGA GATCTCTG AGTACTTGG GGGGTCAGC 60
 CTCTTCCGAG AACAACCTCT GAAGATCTCT GCGGCGGAAA GAGGTGCT GA TTGACGCTC 120
 TATGACGGGT TGAACGACGA GATGCTCGG GATATGATA TCGAAGCTT GGGCTAACAG 180
 GTTCGCAAGA TTTGCGAGT TTATCTCTCG GATGAGCTCT AGGATATG GTTTTCGGAC 240
 GCGCGGGTGA TTTTCTGGA GAGGAGCTT TCGAGAGCG AGTATCTAT AGAGAGGCTG 300
 GACGGCATCA GGTCTCTTG GCGGCGGAG ATGAAGCTT GGTATCTT TCAGATCTC 360
 AAATCTGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA 420
 ATGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA 480
 TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA 540
 TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA 600
 TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA 660
 TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA 720
 TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA 780
 TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA 840
 TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA 900
 TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA 960
 TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA 1020
 TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA 1080
 TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA TTTGATGA 1140

```

TTCCTACAGC AGGATTCAGC AGGCGACGCA ACCTCTGGT GCGAATTTC CAGGAGCTCC      900
CTTGAAACCG CTGGCCCGCT GCTCACTGCC CTACGTAAT CCGCTGCCCT CAGGCCGACC      960
CGCCGGCCGA ATACCAGCAG ATCCGACAGC GAATTGCCGC CCAGCCGGTT GGAGCCCTGC      1020
ATACGCCCGG CACACTCACC GGCAGCGAAC AGGCTGGCA CCGTGGCGG GCGGTGTCC      1080
CGGTCTACTT CGACACCGCC CATCACGTAG TGACACGTG CCCCAGCTT CATTCCTGC      1140
GTTCCGCCAG AG                                     1150

```

(12) INFORMATION FOR SEQ ID NO:141:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 656 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(16) SEQUENCE DESCRIPTION: SEQ ID NO:141:

```

CTCTGCGGCA TTGGGCGAGG TGTACTTGGC GGTCTGTAT GCGGATGAG TCCGACGAC      60
GAGCATGCGC GCAACAGGAC GGTATCGGCT CAGGATGCT ATTGGTATA CTGCGCGAT      120
CGGTGAGT CCGCCCTGCG GGGCTCTTTC CTGGGCGAG GTGATCGAT GTGCTTCGGC      180
CTGCTTTCG CCGCATGAG GCGGCGCGCC GGTGCGGCG GCGCTATTC TGAAGTGG      240
ATGATGACAG CAGAGATMG GGTCTTTTGA AGCTTTTGA GGTCTGCGC GCGATCTTC      300
TCTGCAATT CTACTGGA GAAATCTTC CAGATAGCA TTTGAGGAA TTGGTGGG      360
TGCAGCGGAT GTTCATTCCT CATGCGCGCG AGGATGCGT CTGACGAT GTGCTTCGGC      420
TTGCGGAGG TACGCTTA GGTATGAA TTTCTTAA GTTCTTTC TTGAGAGG      480
TCTGAGCTT ATTAATTT TTTGAGAA ATTTCTTC TTTCTTTC TTTGAGAG      540
TCTGAGCTT ATTAATTT TTTGAGAA ATTTCTTC TTTCTTTC TTTGAGAG      600
TCTGAGCTT ATTAATTT TTTGAGAA ATTTCTTC TTTCTTTC TTTGAGAG      660

```

(17) INFORMATION FOR SEQ ID NO:142:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 656 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:142

Asn Ala Val Val Ala Phe Ala Val Ile Gly Phe Ala Ser Leu Ala Val
 1 5 10 15
 Ala Val Ala Val Thr Ile Arg Pro Thr Ala Ala Ser Lys Pro Val Glu
 20 25 30
 Gly His Gln Asn Ala Gln Pro Gly Lys Phe Met Pro Leu Leu Pro Thr
 35 40 45
 Gln Gln Gln Ala Pro Val Pro Pro Pro Pro Pro Asp Asp Pro Thr Ala
 50 55 60
 Gly Phe Gln Gly Gly Thr Ile Pro Ala Val Gln Asp Val Val Pro Arg
 65 70 75 80
 Pro Gly Thr Ser Pro Gly Val Gly Gly Thr Pro Ala Ser Pro Ala Pro
 85 90 95
 Glu Ala Pro Ala Val Pro Gly Val Val Pro Ala Pro Val Pro Ile Pro
 100 105 110
 Val Pro Ile Ile Ile Pro Phe Phe Pro Gly Trp Gln Pro Gly Met Pro
 115 120 125
 Thr Ile Pro Thr Ala Pro Pro Thr Thr Pro Val Thr Thr Ser Ala Thr
 130 135 140
 Thr Pro Pro Thr Thr Phe Phe Thr Thr Pro Val Thr Thr Pro Pro Thr
 145 150 155 160
 Thr Pro Pro Thr Thr Phe Val Thr Thr Phe Thr Thr Pro Pro Thr
 165 170 175
 Thr Pro Val Thr Thr Phe Phe Thr Thr Val Ala Pro Thr Thr Val Ala
 180 185 190
 Thr Thr Thr Val Ala Thr Thr Thr Val Ala Thr Thr Thr Val Ala Thr
 195 200 205 210
 Ala Thr Ala Thr Thr Thr Thr Val Ala Thr Thr Thr Thr Thr Thr Thr
 215 220 225 230 235 240
 Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
 245 250 255 260 265 270 275 280 285 290 295 300

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 174 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

```

Ile Asn Gln Pro Leu Ala Pro Pro Ala Pro Pro Asp Pro Pro Ser Pro
1           5           10           15
Pro Arg Pro Pro Val Pro Pro Val Pro Pro Leu Pro Pro Ser Pro Pro
20           25           30
Ser Pro Pro Thr Gly Trp Val Pro Arg Ala Leu Leu Pro Pro Trp Leu
35           40           45
Ala Gly Thr Pro Pro Ala Pro Pro Val Pro Pro Met Ala Pro Leu Pro
50           55           60
Pro Ala Ala Pro Leu Pro Pro Leu Pro Pro Leu Pro Pro Leu Pro TH.
65           70           75           80
Ser His Pro Pro Arg Pro Pro Ala Pro Pro Ala Pro Pro Ala Pro Pro
85           90           95
Ala Cys Pro Phe Val Pro Val Pro Pro Ala Pro Pro Leu Pro Pro Ser
100          105          110
Pro Pro Thr Thr Leu Pro Ala Asp Ala Ala Cys Pro Pro Ala Pro Pro
115          120          125
Ala Pro Pro Leu Ala Pro Pro Ser Pro Pro Ala Gly Ser Ala Ala Thr
130          135          140
Arg Ala Leu Thr Gly Ala Thr Ser Val Thr Thr Leu Thr Thr Ala Ala
145          150          155          160          165          170
Leu Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
175          180          185          190          195          200

```

2. INFORMATION FOR SEQ. ID NO: 144:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 174 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

146

Gln Pro Pro Ala Glu Val Ser Asp Gln Arg Val Ser Gly Leu Thr Gly
1 5 10 15

Ala Val Gln Pro Ser Pro Arg Thr Thr Ala Glu Asp Pro Arg Pro Arg
20 25 30

Asn Arg Arg
35

(2) INFORMATION FOR SEQ ID NO:145:

1. SEQUENCE CHARACTERISTICS:

(A) LENGTH: 104 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

2. MOLECULE TYPE: peptide

(3) SEQUENCE DESCRIPTION: SEQ ID NO:145:

Arg Ala Asp Ser Ala Gly Cys Thr Cys Arg Trp Cys Xaa Pro His Glu
1 5 10 15

Cys Arg Arg Pro Ala Met Arg Gln Gln His Gly Ser Arg Ser Thr Thr
20 25 30

Arg Pro Gly Pro Arg Gly Arg Ser Ala Arg Val Arg Pro Gly Arg Leu
35 40 45

Phe Pro Trp Ala Gly Ser Ser Asp Val Phe Pro Pro Trp Phe Ala Ala
50 55 60

Glu Met Trp Ala Arg Arg Val Gly Arg Pro Val Trp Pro Xaa Val Asp
65 70 75 80

Gln His Thr Arg Asp Thr Gly Leu Cys Cys Leu Phe Gln Arg Arg Ala
85 90 95

100

3. INFORMATION FOR SEQ ID NO:146:

1. SEQUENCE CHARACTERISTICS:

(A) LENGTH: 51 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

147

(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

GGATCCATAT GGGCCATCAT CATCATCATC ACGTGATCGA CATCATCGGG ACC

53

(2) INFORMATION FOR SEQ ID NO:147:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: (desc) - "PCR Primer"

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

CCCGAATTCA GGGCTCGCTT GCGCCGGCCT CATCTGAAC GA

42

(2) INFORMATION FOR SEQ ID NO:148:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: (desc) - "PCR Primer"

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:148:

GATGTTTCA GGTGAATG GATGATGTTT

(2) INFORMATION FOR SEQ ID NO:149:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:149:

CTCTGAATTC AGCGGTGGAA ATCGTCGCCA T

31

(2) INFORMATION FOR SEQ ID NO:150:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: (desc) "PCR primer"

(iii) ORIGINAL SOURCE:

(A) ORGANISM: *Mycobacterium tuberculosis*

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:150:

GGATCCAGCG CTGAGATGAA GAGGATGCG GT

33

(2) INFORMATION FOR SEQ ID NO:151:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: (desc) "PCR primer"

(iii) ORIGINAL SOURCE:

(A) ORGANISM: *Mycobacterium tuberculosis*

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:151:

AGAGAGATT TTA GAG GAG ATTTCAGG GAA

(2) INFORMATION FOR SEQ ID NO:152:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:152.

TGTTCTTTCGA	CGGCAGGCTG	GTGCAGCAAG	GGCCCCACCGA	ACAGCTGTTT	TCCTCGCCGA	60
AGCATGCGGA	AACGSECCGA	TACGTGCGCC	GACTGTGCGG	GGACGTCAAG	GACGCCAAGC	120
GGGGAATTTG	AAGAGACAG	AAAGGTATGS	C GTG AAA ATT	CGT TTG CAT ACG		170
			Val Lys Ile Arg	Leu His Thr		
			1	5		
CTG TTG GCG	GTG TTG AGC	GCT GCG GCG	CTG CTG CTA	GCA GCG GCG	GCG	220
Leu Leu Ala	Val Leu Thr	Ala Ala Pro	Leu Leu Leu	Ala Ala Ala	Gly	
	10	11		20		
TGT GCG TCG	AAA GCA GCG	AGC GGT TGG	GTT GAA ACG	GCG GCG GCG	GCG	260
Cys Gly Ser	Lys Pro Pro	Ser Gly Ser	Pro Glu Thr	Gly Ala Gly	Ala	
	25	30		35		
GGT ACT GTC	GCG ACT ACC	GCA GCG TCG	TCG CCG GTC	ACG TTG GCG	GAG	310
Gly Thr Val	Ala Thr Thr	Pro Ala Ser	Ser Pro Val	Thr Leu Ala	Glu	
	40	45		50	55	
ACC GGT ACC	ACG CTS	CTC TAA	CGT CTG TTT	AAC CTG TCG	GGT CCG GCG	360
Thr Gly Ser	Thr Leu Leu	Tyr Pro Leu	Phe Asn Leu	Tyr Gly Pro	Ala	
	60		65		70	
TTT GAG GAG	AGG TAT GCG	AAI GTC AUG	ATC ACC GGT	GAG GCG AGC	GGT	410
Phe His Glu	Arg Tyr Pro	Asn Val Thr	Ile Thr Ala	Gln Gly Thr	Gly	
	75		80		85	
TTT GGT GCG	GCA ATC GAT	AAI GAT GAT	GAT GAT GAT	ATC GAT AAC	ATT GTC	460
Ser Gly Ala	Gly Ile Ala	Gln Ala Ala	Ala Gly Thr	Val Asn Ile	Gly	
	90		95		100	
TTT TTT AAC	GCG TAT CTT	TTT TTT TAA	GCT TAT ATC	GTC TAT TAT	AAI GAT	500
Ala Ser Asn	Ala Tyr Leu	Leu Leu Leu	Ile Ile Asn	Met Ala Ala	His Lys Gly	
	105		110		115	
TTT GAT AAT	AAI GAT GAT	TTT TTT TTT	TTT TTT TTT	TTT TTT TTT	TTT TTT TTT	540
Leu Val Asn	Ile Ala Leu	Ala Ala Ala	Ala Ala Ala	Ala Ala Ala	Ala Ala Ala	
	120		125		130	
TTT TTT TTA	GCT AAT GAT	TTT TTT TTT	TTT TTT TTT	TTT TTT TTT	TTT TTT TTT	580
Leu Phe Gly	Val Ser Ala His	Leu Tyr Leu	Ala Asn Gly	Leu Val Val	Leu Ala	
	135		140		145	

150

170	175	180	
CAC CCG TCC GAC GGG TCC GGT GAC ACC TTC TTG TTC ACC CAG TAC CTG			748
His Arg Ser Asp Gly Ser Gly Asp Thr Phe Leu Phe Thr Gln Tyr Leu			
185	190	195	
TCC AAG CAA GAT CCC GAG GGC TGG GGC AAG TCC CCC GGC TTC GGC ACC			796
Ser Lys Gln Asp Pro Glu Gly Trp Gly Lys Ser Pro Glv Phe Gly Thr			
200	205	210	215
ACC CTC GAC TTC CCG GCG GTG CCG GGT GCG CTG GGT GAG AAC GCG AAC			844
Thr Val Asp Phe Pro Ala Val Pro Gly Ala Leu Gly Glu Asn Gly Asn			
220	225	230	
GGT GGC ATG GTG ACC GGT TGT GTC GAG AIA TCG GCG TCC GTG GCG TAT			892
Gly Gly Met Val Thr Gly Cys Ala Glu Thr Pro Gly Lys Val Ala Tyr			
235	240	245	
ATG GCG ATC ACC TTC CTC GAG CAG GCG ATT CAA CCG GGA CTC GCG GAG			940
Ile Gly Ile Ser Phe Leu Asp Gln Ala Ser Gln Arg Gly Leu Gly Glu			
250	255	260	
GCG CAA CTA GGC AAT AGC TCT GCG AAT TTC TTG TTG CCG GAC GCG CAA			988
Ala Gln Leu Gly Asn Ser Ser Gly Asn Phe Leu Leu Pro Asp Ala Gln			
265	270	275	
ACC ATT CAG GCG GCG GCG GGT GCG TTC GCA TCG AAA ACC GCG GCG AAC			1036
Ser Ile Gln Ala Ala Ala Ala Gly Phe Ala Ser Lys Thr Pro Ala Asn			
280	285	290	295
TAG CTC ATT TCG ATG ATG CAG CAG GGT GCG TCG GAC GCG TAC CCG ATC			1084
Ile Ala Ile Ser Met Ile Asp Ile Pro Ala Pro Asn Gly Tyr Pro Ile			
300	305	310	
ATC AAC TAC GAG TAC GCG ATC CTC AAG CAG CCG CAA AAG GAC GCG GCG			1132
Ile Asn Tyr Gln Tyr Ala Ile Val Asn Asn Arg Gln Lys Asp Ala Ala			
315	320	325	
ATC CTC CAG ACC TTC CAG CAG CTC CTC CTC CTC CTC CTC CTC CTC CTC			1180
Thr Ala Ile Thr Leu Glu Ala Phe Leu His Trp Ala Ile Thr Asn Ile			
330	335	340	
AAC AAC CTC CTC CTC CTC CTC CTC CTC CTC CTC CTC CTC CTC CTC CTC			1228
Asn Lys Ala Ser Phe Leu Ala Ile Glu His Phe Ala Ile Leu Ile Pro			
345	350	355	
CTC CTC CTC AAC TTC CTC CTC CTC CTC CTC CTC CTC CTC CTC CTC CTC			1276
Leu Thr Thr Lys Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr			
360	365	370	

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 374 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:153:

[illegible]

130 135 140
 Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile Lys Thr
 145 150 155 160
 Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn Leu Pro
 165 170 175
 Gly Thr Ala Val Val Pro Leu His Arg Ser Asp Gly Ser Gly Asp Thr
 180 185 190
 Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Glu Gly Trp Gly
 195 200 205
 Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val Pro Gly
 210 215 220
 Ala Leu Gly His Asn Gly Asn Gly Gly Met Val Thr Gly Cys Ala Glu
 225 230 235 240
 Thr Pro Gly Cys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp Gln Ala
 245 250 255
 Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser Gly Asn
 260 265 270
 Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala Gly Phe
 275 280 285
 Ala Ser Lys Thr Pro Ala Asn His Ala Ile Ser Met Ile Asp Gly Pro
 290 295 300
 Ala His Asn Gly Tyr Pro Ile Ile Asn Tyr His Tyr Ala Ile Val Asn
 305 310 315 320
 Asn Arg His Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala Phe Leu
 325 330 335
 His Thr Ala Ile Thr Asp Gly Asn Lys Ala Ser His Thr Asp His Val
 340 345 350
 His Leu His Ile Leu Ser His Ala Thr Val His Leu Ser Asp Ala Leu
 355 360 365
 Leu Ala Thr Ile Ser Ser
 370

INFORMATION FOR SEQ. NO. 152

SEQUENCE INFORMATION

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:154:

TGTTCCTCGA CGGCAGGCTG CTGGAGGAAG GGCCACCCGA ACAGCTGTTT TCCTCGCGGA 60
 AGCATCGCGA AACCGCCCGA TACCTCSCCG GACTGTGCGG GGACGTCAAG GACGCCAAGC 120
 GCGGAAATTG AAGAGCACAG AAAGGTATCG CGTGAAATT CGTTTGCATA CGCTGTTGCG 180
 CGTGTGAGC GTTGGCCCGC TCCTGCTAGC AGCGGCGGGC TGTGGGTCCA AACCACCGAG 240
 CGTTGCGCT GAAACGCGCG CCGGCGCGCG TACTGTGCG ACTACCCCG CGTGTGCGC 300
 GGTGACCTG CGGAGACCG GTAGCAGCT GCTGTACCG GTGTCAAGT TGTGGGTGCG 360
 GGGTTTCAG GAGAGGTATG TAAAGTCAG GATCAGCGCT CAGGCGACCG GTTGTGCTG 420
 CGGATCGCG CAGCTGCTG TGGGACGCT CAGCTTGGC CGCTGCGCG CCGTATGCTG 480
 GGAAGGTGAT ATGGCGCGCG ACGAGCGCT GATGAACATC CGCTGCGCG TGTGCGCTG 540
 GAGGTCAAG TACAGCTGCG CCGGATTGAG CGAGCACCTC AAGCTGACCG GAAAGCTGCT 600
 GCGCGCGATG TACAGCGCG CCGTCAAAAC CTGGGACGAG CGGCGAGTGG CTGCGCTGCG 660
 CCGCGCGCTG AAGCTGCGCG GAGCGCGCT AGTTGCGCT CAGCTGCGCG ACGGCTGCG 720
 TGACAGCTG TTGTTGCGCG AGTACCTGCT CAAGCAAGAT GTGAGCGCT GCGGCAAGT 780
 GCGCGCGCTG GCGACGAGCG TGAGCTGCG CCGGCTGCGCG GTGCGCTGCG GTGAGAGCG 840
 CAGCGCGCG ATGCTGCGCG GTTGGCGCG GAGAGCGCG TGCTGCGCT ATATGCGCT 900
 CAGCTGCTG CAGAGCGCG GTGAGCGCG AGTGGCGCG CGTCAAGTAG GAGAGCTG 960
 TGGGATTTG TTGTTGCGCG AGGCGCGCG GATGAGCGCG CGGCGCGCG GTTGGCGCG 1020
 GAAAGCGCTG GCGAGCGCG GGTGTTGCT GATGAGCGCG CGGCGCGCG AGGCGCTG 1080
 GATGATGAG TACAGTATG GTTGTGCG CGTCAAGTAG GAGAGCTG GAGAGCTG 1140
 GAGTGTGAG GTTGTGCGCG AGGCGCGCG GATGAGCGCG CGGCGCGCG GTTGGCGCG 1200
 GAGTGTGAG GTTGTGCGCG AGGCGCGCG GATGAGCGCG CGGCGCGCG GTTGGCGCG 1260
 GAGTGTGAG GTTGTGCGCG AGGCGCGCG GATGAGCGCG CGGCGCGCG GTTGGCGCG 1320
 GAGTGTGAG GTTGTGCGCG AGGCGCGCG GATGAGCGCG CGGCGCGCG GTTGGCGCG 1380
 GAGTGTGAG GTTGTGCGCG AGGCGCGCG GATGAGCGCG CGGCGCGCG GTTGGCGCG 1440

```

TGGCCCTGAT CATCCGCGTG CCGGTCTCTG TAGGAGCGGC GCTGTGATG GTGGAACGGC 1680
TGGCGAAACG GTTGGCCGAG GCTGTGGGAA TACTCTGGA ATTGCTCGGC GGAATCCCCA 1740
GCGTGGTGGT CGGTTTGTGG GGGGCAATGA CGTCCGGGCG GTTCATCGCT CATCACATCG 1800
CTCCGGTGAT CGCTCACAAC GGTCCCGATG TCCCGGTGCT GAACTACTTG CGCGGCCACC 1860
CGGGCAACGG GGAGGGCATG TTGGTGTCCG GTCTGTGTT GCGCGTCATG GTCGTTCCCA 1920
TTATCGCCAC CACCAATCAT GACCTTTTTC GCGAGGTGCT GGTGTTCGCC CGGGAGGGCG 1980
CGATCGGGAA TTC 1993

```

2. INFORMATION FOR SEQ ID NO:155:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 324 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

```

Met Lys Ile Arg Leu His Thr Leu Leu Ala Val Leu Thr Ala Ala Pro
1          5          10          15
Leu Leu Leu Ala Ala Ala Gly Tyr Gly Ser Lys Pro Pro Ser Gly Ser
20          25          30
Pro Glu Thr Gly Ala Val Ala Gly Thr Val Ala Thr Thr Pro Ala Ser
35          40          45
Ser Pro Val Thr Leu Ala Ile Thr His Leu Thr Leu Leu Tyr Pro Leu
50          55          60
Pro Asn Leu Thr Gly Ile Ala Ile His Val Asn Thr Ile Asn Val Thr
65          70          75
Thr Thr Val Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
80          85          90
Ala Gly Thr Val Asn Leu Gly Ala Ser Ala Ala Ile Leu Ser Leu Gly
95          100          105
Asp Met Ala Ala His Gly Ile Leu Met Asn Thr Thr Thr Ala Ile Leu
110          115          120

```


155

Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn Leu Pro
 165 170 175

Gly Thr Ala Val Val Pro Leu His Arg Ser Asp Gly Ser Gly Asp Thr
 180 185 190

Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Glu Gly Trp Gly
 195 200 205

Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val Pro Gly
 210 215 220

Ala Leu Gly Glu Asn Gly Asn Gly Gly Met Val Thr Gly Cys Ala Glu
 225 230 235 240

Thr Pro Gly Cys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp Gln Ala
 245 250 255

Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser Gly Asn
 260 265 270

Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala Gly Phe
 275 280 285

Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp Gly Pro
 290 295 300

Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile Val Asn
 305 310 315 320

Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala Phe Leu
 325 330 335

His Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp Gln Val
 340 345 350

His Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp Ala Leu
 355 360 365

1. A. The Ile Ser Ser
 370

SEQUENCE INFORMATION

SEQUENCE CHARACTERISTICS
 A. LENGTH 365 amino acids
 B. TYPE single chain
 C. TRANSDUCIBLE NO
 D. OTHER INFORMATION

GTAGCGAAAC GSCAACGGGG CCGCGTGGG CAGCTTCTTC AGCGGTGATT TGCAGAACAC 180
 CTCGTGGAAG GTGATGCGGT CGAATTGTGG CCGCGAACG CTGCGGACCA GGCGGATCGG 240
 CTGCAACCCG GCAGCGCCCG TCGTCAACGG GCATCCCGTT CACCGCGACG GCTTGCCGGG 300
 CCCAACGGAT ACCATTATTC GAACAACGGT TGTATCTTT GTCAACGCTG GCCGTACCG 360
 AGCGCGCAC AGGATGTGAT ATGCCATCTG TCGCCGACA GACAGGAGCG AGGCTTATG 420
 ACAGGATTCG GCGTCGAGCC CTACGGGAG CCGAAGTAG TAGAATGCG CCGGAAGCGG 480
 ATGGGCTATA TCGACGAAG CAAGGCTAG GATGCTGT TCGAGGCG CAACCGCACG 540
 TGTCTTATCT TGTGGGAAA CATCATGCG CACTTGGAG GCGTGGGCG GCTGCTGGCG 600
 TCGGATCTGA TCGGATGGG CCGCTCGAG AAGTGAAGG GATCGGAGC GAGCGCTAT 660
 AGCTATGGG AGCAACGAG CTTTTGCTT GCGCTTGGG ATGCGCTCGA CCGCGCGAG 720
 CAGGTGTGAC TGGTGTGCA CCACTGGGG TCGCGCTCG GCTTGCATG GCGTAACGAG 780
 GATCGGAGG GAGTGGGCG GATCGCTTC ATGAGAGGA TCGTCAGCG GATGAGCTG 840
 GCGGATGGG CCGCGCGCT TCGGCTCTT TCGAGCTT TCGGATCGG TCAAGGCGAG 900
 GCAATGGCT TGGAGGAAA CATCTTTTC GATGGCTT TCGGCTCTT GATCTTGGG 960
 CAGCTCAGG AGGAGGAAT GAACGATAT CCGCGCGAT TCGTGAAGG CCGCGAGGAG 1020
 GCTGCGGGA GCTTGTCTG GCGAGGAAAG CTTCAATAG AGCTGAGG GCGGAGCTG 1080
 TCGGCTCTG TCAAGGAGT CCGGATCTT TCGGAGGA AGGATCTT GAGGCTCTT 1140
 ATGAGGCTG AGCGCGGCG GATATCTG GCGGCTCTT GCGGCTCTT GCGGCTCTT 1200
 GCGGCTCTT GCGGCTCTT GCGGCTCTT GCGGCTCTT GCGGCTCTT GCGGCTCTT 1260
 GCGGCTCTT GCGGCTCTT GCGGCTCTT GCGGCTCTT GCGGCTCTT GCGGCTCTT 1320
 GCGGCTCTT GCGGCTCTT GCGGCTCTT GCGGCTCTT GCGGCTCTT GCGGCTCTT 1380
 GCGGCTCTT GCGGCTCTT GCGGCTCTT GCGGCTCTT GCGGCTCTT GCGGCTCTT 1440
 GCGGCTCTT GCGGCTCTT GCGGCTCTT GCGGCTCTT GCGGCTCTT GCGGCTCTT 1500
 GCGGCTCTT GCGGCTCTT GCGGCTCTT GCGGCTCTT GCGGCTCTT GCGGCTCTT 1560

TACGTGCGCG TTCACGATTC GGGGGGGGGG TATGCTT

1177

(2) INFORMATION FOR SEQ ID NO:157:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 324 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO:157:

GAGATTGAAT GGTACCGCTC TCCTTAGCGG CTCCTTCCCG TGAATGCGA TATCAGCGAC	60
GGCATGTTT TCCTGTGCA GCTTGGCTCC ATGCTTGGAC GTTGTAAAT GCAGGGTTTG	120
ATCAGTAATT GCGGGGAGC TTCTCGGAA GCGGGGAGG ATGTGCTGA GCGGGGCGCT	180
AGGTTGCGG GAGCGGAGC CTGATGCTC AGTGGCGTC GGTGACTTA GCGAGGTTT	240
GGGCTGCTC GTGACACTG GTACTCCGT GAGGAGCGG GCGGTGCTT GGTGAAGAC	300
GCTGACCGAC GCGGGGATT CAGA	324

(2) INFORMATION FOR SEQ ID NO:158:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1338 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO:158:

ACATATGCG GCGTTTGGG TTTAAGGA GTTATAAT GTAAAGACT TCGCTGCGG	60
AAGGATTGAG GAACGGCTG TGGGCTGT GTATTTCTT AAGGTATGG GTGATGTCT	120
TAATATTTT GTTAAATG TTTATTA TTTTAAAT GTTATTTT GTTAAAGGA	180
TAATATTTT AATTAAT GTTAAAT GTTAAAT GTTAAAT GTTAAAT	240
TAATATTTT GTTAAAT GTTAAAT GTTAAAT GTTAAAT GTTAAAT	300
TAATATTTT GTTAAAT GTTAAAT GTTAAAT GTTAAAT GTTAAAT	360
TAATATTTT GTTAAAT GTTAAAT GTTAAAT GTTAAAT GTTAAAT	420
TAATATTTT GTTAAAT GTTAAAT GTTAAAT GTTAAAT GTTAAAT	480

GGAGCCCGCCG GGTGACCGCG GTGCTGGACG GCGGCGAGAC GGTGCTGCTG CGTAAGGGCG	720
GGATCGGGGA GAAGCCCTTC GAGGTGGCGG CCCACGAGTT CTTGTTGTTT CCGACGGTCG	780
CGCACAGCCA CCCCAGCGCG GTTCGCCCCG AGCACCAGCA CCTGCTGGGC CCGGCGGCCG	840
CCGACAGCAC CGACGAGTGT GTGCTACTGC GGGCCGCGAG GAAAGTTGTT GCCGCACTGC	900
CGGTTAACCG GCGAGAGCGT CTGGACGCCA TCGAGGATCT GCACATCTGG ACCGCGGAGT	960
CGGTGCGGCG CGACGAGTTC GACTTTGGCG CCAAGCACA ACTGCGCGTC TTGTTGGTCT	1020
CGGCGATCGG GGTGCTGAGG CCGGTGCGT TGGCGCTAG GCGCGAGTAC GCGGTTGCA	1080
CGAGCTGGGT GAGGCTGCGG GTGACGCGGA CGTTGCGCG CCGGTGCGA GAGAGCGCG	1140
CGGTGCGCGA GGTGCTGAGG CCGGTGCGT AGGCTGCGT TCGAGTGGCG GCGATCGCTT	1200
GGGTGCGAGG GTGACGCGGA GTGCGGCGT GAGTGTCTCT GGTGCGCTT CCGTCTGCG	1260
TGCGGTCAAT TGACGCGCGG GCGACAGCA GATTGCGCG CCGATCTCT CCGCGCGCGG	1320
GCGCGAGCG GTAGAGCG	1338

(2) INFORMATION FOR SEQ ID NO:159:

(A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 321 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(B) SEQUENCE DESCRIPTION: SEQ ID NO:159:

CGGTGCGCGA GGTGCTGAGG CCGGTGCGT GCGGCGAGCA GGTGCTGCTG CGTAAGGGCG	60
TGGGTGCGCG GCGGAGCGCG GAGGCTGCGT TCGGTGCGCA GAAAGTTGTT AACGCGGAG	120
CGGTGCGCGA GGTGCTGAGG CCGGTGCGT AGGCTGCGT TCGAGTGGCG GCGATCGCTT	180
GGGTGCGAGG GTGACGCGGA GTGCGGCGT GAGTGTCTCT GGTGCGCTT CCGTCTGCG	240
TGCGGTCAAT TGACGCGCGG GCGACAGCA GATTGCGCG CCGATCTCT CCGCGCGCGG	300
GCGCGAGCG GTAGAGCG	321

(2) INFORMATION FOR SEQ ID NO:160:

(A) SEQUENCE CHARACTERISTICS:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:160:

```

GAAGACCCCGG CCCCCCATA TGGATCGGTT CGCCGACTAC TTTCGGCGAA CGTGCACGGG      60
GCGGCGTCCG GCTGATCATC ACGGGTGGCT ACGCGCCCAA CGGCACCGGA TGGCTGCTGC      120
CGTTCGCCTC CGAACTCGTC ACTTCGGGGC AAGCCCCGACG GCACCGCCGA ATCACCAGGG      180
CGGTCCACGA TTGGGGTGCA AAGATCCTGC TGCAAATCCT GCACCGCGGA CGGTACCCCT      240
ACCACCCACT TGCSTCAGC GCCTCGCGGA TCAAGGCGCC GATCACCUCG TTTCGTCCCG      300
GAGCACTATC GCTCGCGGG GTGAAGCGA CCATCGCGGA TTTCGCGCG TGGGGGCACT      360
TGGCGCGCGA TGGGGGTAC GAGGGGTG AATATATGG GAGCGAAGG TATCTGCTCA      420
ATCAGTTCCT GCGCGCGCGC ACAAACGAG GCACCGACTC TGGGGTGGG AAGACGCGCA      480
ACCTTCGCGG GT

```

(2) INFORMATION FOR SEQ ID NO:161:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 536 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:161:

```

Phe Ala Gln His Leu Val Glu Gly Asp Ala Val Gln Leu Trp Arg Ala      1
1 5 10 15
Asn Ala Ala Asp Gln Ala Asp Arg Leu Gln Pro Gly Ser Ala Arg Arg      20
20 25 30
Gln Arg Ala Ser Ala Ser Pro Arg Arg Leu Ala Gly Pro Asn Ala Tyr      35
35 40 45
His Tyr Ser Asn Asn Arg Arg Thr Leu Val His Arg Thr Trp Leu Trp      50
50 55 60
Pro Ala Ala Leu Ala Val Val Val Val Val Val Val Val Val Val Val      65
65 70 75 80 85 90 95 100 105 110 115 120
Pro Gly Leu Met His Ala His Gly Leu Val Val Val Val Val Val Val      125
125 130 135 140 145 150 155 160 165 170 175 180
Leu Val Val Val Val Val Val Val Val Val Val Val Val Val Val Val      185
185 190 195 200 205 210 215 220 225 230 235 240

```


161

Gln Arg Ser Ser Tyr Leu Pro Ser Glu Leu Val Ala Ala Phe Leu Trp
 435 440 445

Ala Gln Phe Glu Glu Ala Glu Arg Ile Thr Arg Ile Arg Leu Asp Leu
 450 455 460

Trp Asn Arg Tyr His Glu Ser Phe Glu Ser Leu Glu Gln Arg Gly Leu
 465 470 475 480

Leu Arg Arg Pro Ile Ile Pro Gln Gly Cys Ser His Asn Ala His Met
 485 490 495

Tyr Tyr Val Leu Leu Ala Pro Ser Ala Asp Arg Glu Glu Val Leu Ala
 500 505 510

Arg Leu Thr Ser Glu Gly Ile Gly Ala Val Phe His Tyr Val Pro Leu
 515 520 525

His Asp Ser Pro Ala Gly Arg Arg
 530 535

(2) INFORMATION FOR SEQ ID NO:162:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 294 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

X1 SEQUENCE DESCRIPTION: SEQ ID NO:162:

Asn Glu Ser Ala Pro Arg Ser Pro Met Leu Pro Ser Ala Arg Pro Arg
 1 5 10 15

Tyr Asp Ala Ile Ala Val Leu Leu Asn Glu Met His Ala Gly His Cys
 20 25 30

Asp Phe Gly Leu Val Gly Arg Ala Thr Asp Ile Val Thr Asp Ala Ala
 35 40 45

His Asp Asp Arg Ala Gly Leu Gly Val Asp Glu Glu Phe Arg His Val
 50 55 60

Gly Phe Leu Glu Pro Ala Pro Val Leu Val Asp His Arg Asp Asp Leu
 65 70 75 80

Gly Gly Leu Thr Val Asp Thr Lys Val Ser Trp Pro Arg Gln Arg Gly
 85 90 95

Asp Lys Asp Val Val Leu Gln Arg His Trp Leu Ala Leu Arg Arg Ser
 130 135 140
 Glu Thr Leu Glu His Thr Pro His Gly Arg Arg Pro Val Arg Pro Arg
 145 150 155 160
 His Arg Gly Asp Asp Arg Phe His Glu Arg Asp Pro Leu His Ser Val
 165 170 175
 Ala Met Leu Val Ser Pro Val Glu Ala Glu Arg Arg Ala Pro Val Val
 180 185 190
 Gln His Gln Tyr His Val Val Ala Glu Val Glu Arg Ile Pro Glu Arg
 195 200 205
 Glu Gln Lys Val Ser Leu Leu Ala Ile Ala Ile Ala Val Gly Ser Arg
 210 215 220
 Trp Ala Glu Leu Val Arg Arg Ala His Pro Asp Gln Ile Ala Gly His
 225 230 235 240
 Gln Pro Ala Gln Pro Phe Gln Val Arg His Asp Val Ala Pro Gln Val
 245 250 255
 Arg Arg Arg Gly Val Ala Val Leu Lys Asp Asp Gly Val Thr Leu Ala
 260 265 270
 Phe Val Asp Ile Arg His Ala Leu Pro Gly Asp Phe
 275 280

3. INFORMATION FOR SEQ ID NO:163:

1. SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 264 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:163

ATGAACATCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT
 1
 GCGATGCGAG CGATGCGAG CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT
 11
 ATGCGCGCTCT GCGTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT
 21
 GCGCGCGAG CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT
 31
 GCGCGCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT
 41

(B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:164:

TAGTCGGCGA CGATGACGTC GCGGTCCAGG CCGACCGCTT CAAGCACCAC CGCGACCACG	60
AAGCCGGTGC GATCCTTACC CGCGAACCAG TGGGTGAGCA CCGGGCGTCC GCGGGCAAGC	120
AGTGTGACGA CACGATGTAG CGCGCGCTGT GCTCCATTGC GCGTTGGGAA TTGGCGATAC	180
TGCTCGGTCA TGTAGCGGGT GCGCGCGTCA TTTATCGACT GGCTGGATTC GCGCGACTCG	240
CGCTTGGACC CGTCATTGGT TAGCAGCCTC TTGAATGCGG TTTCGTGCGG CGCTGACTCG	300
TGGCGCTCAT CATCGGCGAG CTCGGGGAAC GCGAGCAGGT GGACGTCGAT GCGCTCCGGA	360
AGCGCTCCTG GACCGCGCGG GCGAAGCTCC CCGGACGACC GCAGGTCGGC AACGTCGGTG	420
ATCGCGACGC GCGGCACGCT TGCGCTTCTT GCGGAATTCC GCACGAGGCT GCGGAGCCAC	480
CGCGCATCAC CAAGCAACGC TTGCGCAGTA CGGATCGTCA CTTCCGCATC CGGCAGACCA	540
ATCTCGTCCG CCGCCATCCT CAGATCCCGC TCGTCCCTTG ACAAGAACGG CCGCAGATGT	600
CGCAGCGGCT ATCGGAGATT GAACCGGCA CCGAGTTCTT CAATCGCTGC GCGCTCCCGC	660
AGTATTGCA GTTTCCGCGG CTCGCGGTAT TTAGCAAGCA TCGAGTCTC GACGAACTCG	720
CGCCACGTAA CCGACGGGCT AGCTCCCGGC CTGAGGCGGA CGATCGGCGG GTGATCTTTG	780
CGCGAGCTCT CTAGCGGTT GATCCACCGC TTCTCGGTTC GCGCGCGGAG CGCGATCAGC	840
TTATCGACCT CGGCTATGC CGACCGCAAG CTCGCGCGCT TCGTCGAGGT CAAGAATCTC	900
AGCATCGGCA CGGCGACCAA GGTGCGCGAC CTGAGCTATC TCGCGACGC CGACATCGGC	960
GATACAGCA ACATCGGCTT TCGAGCTT TTTCTAAAT AGGAGGTAC CTCGAAACCG	1020
TCGAGACCG TCGGTTCGA CTAGCGAGT GGTTCGAGA CATTTCTT TCGCGACTA	1080
AGATGCTG ACGGCGCTA TCGCGGCTT GCGAGCTG TCGAGAGGA TCTCGCGCGG	1140
CGCGCGCTG CAGTGTGCGG GGTTCGCAA	1200

1 INFORMATION FOR SEQ ID NO:164:

1. SEQUENCE CHARACTERISTICS

A. LENGTH: 1200 base pairs

B. TYPE: DNA

GCAAAGGCGG CACCGGCGGG GCCGGCATGA ACAGCCTCGA CCGGCTGCTA GCGGCCCAAG 60
 ACGGCGGCCA AGGCGGCACC GCGGCGACCG GCGGCAACGC CGGCGCGGGC GGCACCAAGT 120
 TCACCCAAGG CGCCGACGGC AACGCCGGCA ACGGCGGTGA CGGCGGGGTC GCGGCAACG 180
 GCGGAAACGG CGGAAACGGC GCAGACAACA CCACCACCGC GCGCGCC 227

(2) INFORMATION FOR SEQ ID NO:166:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 304 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:166:

GTGCGCCACC ATGCGCGGGG AGGCGGGTAG CGGTGCGGCC GGCTCTACCC CAGGCGCCAA 60
 GGGCGCCCAAC GGCTTCACTC CAACCAGCGG CCGGCGACGG GCGGACGGCG GCAACGGCGG 120
 CAACTCCCAA GTGCTCGGCG GCAACGGCGG CCGGCGCGGC AATGGCGGCA ACGGCGGCAG 180
 GCGGCGCACG GCGGCGAACC GCGGCGCGGG TGGCGACGGC GCGTTTGGTG GCATGAGTGC 240
 CAACGCGACC AACGCTGGTG AAAACGGGGC AAACGCTAAC CCGGCGCGCA ACGGTGGCGC 300
 CGGC 304

(2) INFORMATION FOR SEQ ID NO:167:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1439 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:167:

TTTGAAGCTT TGGGAGGCTT TATAAGCAAG AATACATGCA TGAAGTCTT TGGGTGAGC 60
 TATGAGCTT ATTTAACAAT TGGGCTTCA CCGGCGAGGG CGAGCATTTG GCGGCGGATC 120
 TATGCGCTGA GGTCTAGGAA TATTCTTTCG CCAATTTGCG TCGGCGGAAA GCGAAGCGGG 180
 GTGCGAGTT GTTTAGCGCG CCGAGCTTG TCAAGCTGAT CCGGAGCTG CTGAAGTCT 240
 TGAATGAGC GTGCTATGAT TGTCTCTG TTTGCGAGG TGTCTCTG TGTCTCTG 300

GGCTCGGGCG CCGATGGAGT GATACCTTCG CCGCGGACCA GCACCCGGAC GTGCAGATGG 480
 ACTACGTGAT GGCCAATCCG CCGTTCACCA TCAAAGACTG GCGCCGCAAC GAGGAAGACC 540
 CACGCTGGCG CTTCGGTGTG CCGCCCGCCA ATAACGCCAA CTACGCATGG ATTACACACA 600
 TCCTGTACAA CTTGGCGCCG GGAGGTCCGG CGGGCGTGGT GATGGCCAAC GGGTCGATGT 660
 CGTCCAACTC CAACGGCAAG GGGGATATTC GCGCGCAAAT CGTGGAGGCG GATTTGGTTT 720
 CCGTCATGGT CCGCTTACCC ACCCAGCTGT TCCGCAGCAC CCGAATCCCG GTGTCCCTGT 780
 GGTTCCTTGC CAAAAACAAG GCGGCAGGTA AGCAAGGGTC TATCAACCGG TCGGGGACG 840
 TCGTGTTCAT CGACGCTCGT GAACTGGGCG ACCTAGTGGG CCGGGCCGAG CCGGCGCTGA 900
 CCAACGAGGA GATCGTCCCG ATCGGGGATA CTTTCACCGC GAGCAGGACC ACCGCGAACG 960
 CCGGCTCCCG TCGTCCCGCG GGTAAATGGG GCACTGGGCT CAACGGCGCG GCGGCTGCTC 1020
 GCGGGGCGCG CCGCAACCGG GGTGTGCGCG GCGTGTCTTT CCGCAACGCT GTGGGCGGCG 1080
 ACGGGCGCAA CCGCGGCAAC GCGGCGGACG GCGGCGACGG CACGACGGCG GCGGCGGCGG 1140
 GCAAGGCGCG CAACGGCAGC AGGGGTGCGG CAGCGGGCTC AGGCGTCTGC AACGTACCG 1200
 CCGGTCACCG CCGCAACCGG GGCAATGCGG GCAACGGGCG CAACGGCTCC GCGGGCGCGG 1260
 CCGGCGAGCG CGGTCCCGCG GCGAGGCGCG GCAACGGGCG CACGCGCGCG GTTGCCACCG 1320
 CCGGCGCGCG CCGCAAGGCG GGCAACGGCA CAGCGGCTCG CCGGCGCGCG TCAGGCGCTC 1380
 TCAACGTCAC TCGCGGCGCG GCGGCGAACG CCGGCAATCG CCGCAACGCG GCAACGCG 1440

2. INFORMATION FOR SEQ ID NO.168:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 329 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO.168

GCGCGGCGCG GCGCGGATTT TCTGTGCTCT TCAATCTCGT TCGGATAAC GCGGCTGATC 40
 CTGCTAACCG CCGGATGGGT GCGGCTGCGG GCGGTCGCGG CCGGCTGCGG GCGGCGCGCG 100
 TCAACGCTCT TCTGTGCTCT TCAATCTCGT TCGGATAAC GCGGCTGATC 160

(2) INFORMATION FOR SEQ ID NO:169:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 80 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:169:

GCAACGGTGG CAACGGCGGC ACCAGCACGA CCGTGGGGAT GGCCGGAGGT AACTGTGGTG 60
 CCGCCGGGGCT GATCGGCAAC 80

(2) INFORMATION FOR SEQ ID NO:170:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 392 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:170:

JGGGTGTGTC GCACTCACAC CGCGGATTC GCGGACGTTG GCGCGCCAAT ATCCAGGTCA 60
 AGGCGTACTA GTTACCGTCC GAGGACCGCC GCATCAAGGT GCGGGTCAGC GCGCAAGGAA 120
 TCAAGGTGAT CGACCGCCAC GGGCAGCAG GCGGTGCTCC CCGCGCTCCG GCAGGATCCG 180
 CCGCGGGCCA GTTCGCGCGC CAGCGGGGT CATGCTGTC AACCGCGCCG ATCGTGTGAG 240
 TACAATGAT GCGGCGCAG CAGATCGTC CATTTGTA GCGGTGTTCC ACCGCAGGGA 300
 CCGGTATAC GATGTCGAC GATGTCGAC TCAAGAAC GCGATACCA TCGGTGATC 360
 CCGCGACAGC CAGGAGTCC AAGACCGTTA CA 392

(2) INFORMATION FOR SEQ ID NO:171:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 535 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:171:

ATCGGTGTA CCGGCGCA TCGGTGTA TCGGTGTA TCGGTGTA TCGGTGTA TCGGTGTA TCGGTGTA

GGCGGCACCG GCGGCACCGG CGGAGCGGGG GGAGCGGGCG GGGCGGGTGG CGCCATCGGT	240
ACCGGGGGCA CCGGCAGCGC GGTGGGCAGC GTGGTAACG CCGGGATCGG CGGTACCGGC	300
GGTACGGGTG GTGTGGGTGG TGCTGGTGGT GCAGGTGGG CTGCGGGCCG TGGCAGCAGC	360
GCTACCGGTG GCGCCGGGTT CGCGGGCGGC GCCGGCGGAG AAGGCGGACC GGGCGGCAAC	420
AGCGGTGTGG GCGGCACCAA CGGCTCCGGC GCGCGGGCG GTGCAGGCGG CAAGGGCGGC	480
ACCGGAGGTG CCGCGGGGTC CGGCGCGGAC AACCGCACCG GTGCTGGTTT CGCCG	535

(2) INFORMATION FOR SEQ ID NO:172:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 590 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO:172:

CGAAGTGGG GGGGGCGATA GGGGGGTGAC CGACTACTAC ATCATCCGCA CCGAGAATCG	60
GCGGTGCTG CAACGGGTGC GGGGGTGGC GGTTCATCGGA GATCCGGTGG CCGACCTGAT	120
CGAGCGAAG CTGAAGGTGA TCGTCAAGTT GGGGTACGGC GACCGGAAC ACGGCTACTC	180
GACGAGCTAC GCGGATGTGC GAACGGGTT GGGGCTGTGG CCGAACGTGC CGCCTCAGGT	240
GATCGCGGAT GCGGTGGCGG CGGGAACACA AGAAGGCGTC GTTGACTTCA CCGCGGACCT	300
GGAGGGGCTG TCGGGCGAAC GGTTCAGGCT GGTGAGATG GAGCTGGCGC AACCGCGCGA	360
GTGGGTGGCG GCGGTGGCGG CGGACCGGAG GCGGGCGGAG GTGGTGAACA CGCTCGCCAG	420
GATCATCTCA AGCAACTACG CGTCTGTGTT GCGGACGCTG GACATCGCGC TCGCCTGGTC	480
AGCAAGCTGC GGTGTACAG GAGCAACTT TGGTCAAGT AACTGGGTGC GGGCAATCTG	540
ATGAAAGGAA TGGGTATCG GTGGGGGCT GGTTCAGTT TACCAAGAT GATAGCGCG	600
GTGGTGGAA GTGGTGGCG TCGTGGCGG GGTTCAGTT TACCAAGAT GATAGCGCG	660
GTGGTGGAA GTGGTGGCG TCGTGGCGG GGTTCAGTT TACCAAGAT GATAGCGCG	720

(2) INFORMATION FOR SEQ ID NO:173:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 177 base pairs

```

ACGGTGACGG CGGTACTGGC GCGGGGACG GCGGCAACGG CGGGAATCCC GGGTGGCTCT      60
TGGGCACAGC CGGGGGTGGC GGCAACGGTG GCGCCGGCAG CACCGTACT GCAGGTGGCG      120
GCTCTGGGGG CACCGGGGGC GACGGCGGGA CCGCGGGGCG TGGCGGCCCTG TTAATGGGCG      180
CCGGCGCCCG CGGGCACGGT GGCCTGCGG GCGCGGGCGG TGCCGGTGTG GACGGTGGCG      240
GCGCCGGCGG GGCCGGCGGG GCGGGCGGCA ACGGCGGGCG CGGGGGTCAA GCCGCCCTGC      300
TGTTCCGGCG CGGCGGCACC GCGGGAGCCG GCGGCTACGG CGCGGATGGC GGTGGCGGCG      360
GTGACGGCTT CGACGGCACG ATGGCCGGCC TGGGTGGTAC CGGTGGC                      407

```

(2) INFORMATION FOR SEQ ID NO:174:

1. SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 468 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:174:

```

GATCGGTUAG CGCATGCGCC TCGCGGCA GCGATTCCGC GGTCTCACCG AAGAACATCG      60
TGCACCGCGG GCGCGGAGC AGCCCGCTGC GTCGCGGCGC GTCBAACGCC TCCAGCAGGC      120
ACAGCCAGTC GTTGGCGGCT TCGGAGGCGA ACACGTCGGT CTCACCGGTG TAGATCGCGG      180
CGATCGCGGC CTCGCGAAC GCATTGCGGC ACGCGCGGCG GTCTTTGTGA TGTCTGACGA      240
TACGCGCGAT GTCTCGGCGC ACCACGCGC GCGCGCGGAA GTTGGCGGCG CTGCGCGAGTA      300
GCGCGCGGAC CTCGCGCGGC AGGTCTTGG GATCTGCGG CGGAGCGGCT CCGCGCGGAC      360
GCGCGAAAAA CGACCGCTCA CGCATCTGG TCGCTCTGG ATATCGCTTG CGTCTCTGGG      420
GATATTGGA CGCGCATGCG CGCATCTGG TCGCTCTGG ATATCGCTTG CGTCTCTGGG      468

```

(2) INFORMATION FOR SEQ ID NO:175:

1. SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 219 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:175:

GGCGGCGCGG GCGGCAATGG CGGCGCGGGC GGCAACGGCG AGGCGGCGCG GTACACCGAC 180
 GCGGCGCGCG GCACCGCGCG CGACGGCGCG AACGGCGCG 219

(2) INFORMATION FOR SEQ ID NO:176:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 494 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:176:

TAGCTCGCGG GAGGCGCGCA AGGCGCGCGA CGGTGGCCAC GGCGGTGACG GCGTCGGCGG 60
 CAACAGTTCC CTCACCGAAG GCGGCAGCGG CGGTGGCGCG GCGCGCGCGG GCGCGCGCGG 120
 CAGCGGCTTT TTGCGCGCGA AGGCGCGGCTT CGCGCGCGAC GCGCGTCAGG GCGGCCCCAA 180
 GCGCGCGCGT ACCGTGCGCA CGGTGCGCGG TGGCGCGCGG AACGCGCGGTG TCGCGCGCGG 240
 GCGCGCGCGAC GCGCTCTTTG CGGTGCGCGG CGGCGAGCGG GCGCTCGGTG GGCAGGCGCG 300
 CAATGGCGCG GCGTCCACCG GCGGCAACCG CGGCTTTGGC GCGCGCGCGG GTGGCGGAGG 360
 CAACGCGCGG GCGCTCGCGG AATCGCGGCT GAGCATGGAC AGGCGCGCGA AGTTGCGCTG 420
 CATCGCATCA GCGCGCTACT GCGCGCAACA CCGGCAACAT CAGCGGAGTT AGCGGGGCGG 480
 ATTTCTGTAT CACC 494

(2) INFORMATION FOR SEQ ID NO:177:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 220 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:177:

GCGGCGCGG GCGGCGCGG GCGGCGCGG GCGGCGCGG GCGGCGCGG GCGGCGCGG 60
 TGGCGCGCG GCGGCGCGG GCGGCGCGG GCGGCGCGG GCGGCGCGG GCGGCGCGG 120
 TGGCGCGCG GCGGCGCGG GCGGCGCGG GCGGCGCGG GCGGCGCGG GCGGCGCGG 180
 TGGCGCGCG GCGGCGCGG GCGGCGCGG GCGGCGCGG GCGGCGCGG GCGGCGCGG 220

- (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:178:

ATGGCGGCAA CGGGGGCCCC GCGGGTGCTG GCGGGGCGCG CGACTACAAT TTCCAACGGC	60
GGGCAGGGTG GTGCCGCGCG CCAAGGCGGC CAAGGCGGCG TGGGCGGGC AAGCACCACC	120
TGATCGGCCT AGCCGCACCC GGGAAAGCCG ATCCAACAGG CGACGATGCC GCCTTCCTTG	180
CCGCGTTGGA CCAGGCGCGC ATCAGCTAGC CTGACCCAGG CCACGCCATA ACGGCCGCCA	240
AGGCGATGTG TGGGCTGTGT GCTAACGGCG TAACAGGTCT ACAGCTGGTC GCGGACCTGC	300
GGGACTACAA TCGGGGGTGC ACCATGGACA GCGCGGCCAA GTTCGCTGCC ATCGCATCAG	360
GCGGCTACTG GCGCGAACAC CTGGAACA	388

(2) INFORMATION FOR SEQ ID NO:179:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 400 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:179:

GCAAGGGCGG CACCGCGCGG GCGGGCATGA ACAGCCTCGA CCGGCTGCTA GCGGCCCCAAG	60
ACGGGGGGCA AGGCGGCACC GCGGGCAGCG GCGGCAAGCG CCGCGCCCGC GGCACCAAGCT	120
TTACGGAAGG GCGGACGCGC AACCGCGGCA AGGGCGGTGA GCGCGGGGTC GCGCGCAACG	180
GCGAAAGCGG GGGAAACGGC GCAGACAACA GAGCAACCGC GCGCGCCGCG ACCACAGGCG	240
TCGAGCGCGG GCGCGCGCGG GCGGCGGAAA GCGGCGGAAA GCGCGGAGCG GCGGCGACCG	300
TCGCGCGGCG GCGGCGGCGG AACGCGGCGA AGCGGCGA GCGGCGGAAA GCGGCGACCG	360
TCGCGCGGCG GCGGCGGCGG AACGCGGCGA AGCGGCGA GCGGCGGAAA GCGGCGACCG	420

(2) INFORMATION FOR SEQ ID NO:180:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 528 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

171

AGCGGCAACG GCGGCCAAGG GCGGCAGCGG GGGCAACGGC GGCAACGCGG GCATGGGCGG 120
 CAACAGCGGG ACCGGCAGCG GCGACGGCGG TGCGGGCGGG AACGGCGGGC GGGCGGGCAC 180
 GGGCGGCACC GCGGGCGAGC GCGGCCTCAC CGGTACTGGC GGCACCGGGC GCAGCGGTGG 240
 CACCGGCGGT GACGGCGGTA ACGGCGGCAA CGGAGCAGAT AACACCGCAA ACATGACTGC 300
 GCAGJCGGGC GGTGACGJTG GCAACGGCGG CGACGGTGCC TTCGGCGGGC GGGCCGGGGC 360
 CGCGCGCGGT GGCTTGACCG CTGGCGCCAA CGGCACCGGC GGGCAAGGCG GCGCCGGCGG 420
 CGATGGCGGC AACGGGGGCA TCGCGGGCCA CGGGCCACTC ACTGACGACC CCGGCGGGCAA 480
 CCGGGGCACC GCGGCCAAGG GCGGCACCGG CGGCACCGGC GCGCGGGCA TCGGCAGC 538

2. INFORMATION FOR SEQ ID NO:181:

1. SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 339 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:181:

GGGTGGTGG TGCGCGGGG CAGCTCTTCA GCGCGGAGG CGCGCGGGT GCGGTGGGG 60
 TTGGGGGAC GCGGGGCAAG GGTGGGGGTG GCGGTGGCGG AGCGGGCGGC GCGGACGCGG 120
 CCGCGAGGAC AGGTCTTACC GGTGGTACCG GGTGGGTGG GCGGGCGGGC GCGGTGGGG 180
 GCGGGGGCG GAACGGGATT GCGGGCGGCA TGAACGGGTG GGTGGGTGGG GCGGCACT 240

3. INFORMATION FOR SEQ ID NO:182

1. SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 985 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:182

AGCAAGGCTA GCGGTGGGGT GGGGTGGGGT GGTGGGGGGT GGTGAGAAAT GCGAGCGGGT 5
 GCGAGAGCGG GTGTGGGGGG GAGGAGCGG: TCGGGGGGGT GGGGGGGTGG AGCGGGCAAG 11
 GGTGGAGGAG GAGGTGGGGT GGTGGGGTGG GGTGAGAAAT GGTGGGGTGG TGGTTGGGGT 17

GACGGCGGCC AAGGCGGCCA AGGCGGGGCC GGGCGCAGCG CCGCGCCCGG CGGCATCAAC 420
 GGGGCGGGCG GGGCGGGCGG AACCGCGGC GACGGCGGGG ACGGCGCAAC CGGTGCCGCA 480
 GGTCTCGGCG ACAACGGCGG GTCGGCGGT GACGGTGGGG CCGGTGGCGC CGCCGGCAAC 540
 GGGCGCAACG CGGGCGTCGG CTTGACAGCC AAGGCGGGCG ACGGCGGCGG CGCGGGCAAT 600
 GGGCGCAACG GGGGCGCGCG CGGTGCTGGC GGGGCGGGCG ACAACAATTT CAACGGCGGC 660
 CAGGGTGGTG CCGGCGGCCA AGGCGGCCAA GCGGGCTTGG GCGGGGCAAG CACCACCTGA 720
 TCGGCTTAGC CGCACCCTGG AAAGCGGATC CAACAGGCGA CGATGCGCGC TTCCTTGCGG 780
 CGTTGGACCA GGGCGGCATC ACCTACGCTG ACCGCGGCCA CGCATAACG GCGGCGAAGG 840
 CGATGTGTGG GCTGTGTGT AACCGGCTAA CAGGTCTACA GGTGGTGGCG GACCTGCGGG 900
 AATACAATTC CCGGCTGACC ATGGACAGCG CGGCGAAGTT CGCTGCCATC GCATCAGGCG 960
 CGTACTGCGC CGAACACCTG GAACA 985

2. INFORMATION FOR SEQ ID NO:183:

(A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2138 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(B) SEQUENCE DESCRIPTION: SEQ ID NO:183:

CGGCACGAGG ATGGTACCG CCGCGCATCG CAGCTTCTCG ATTGCGCGGG TTTGCGCAGC 50
 CGAGGAAAGC CGCTACGAGA TGGCGCTGCG GAAGTAGGCG GATCGCTTCC CGATGCGGGC 100
 ATGACCGGCG GGCATCAAT TAGTCAGCA ACCTTTCAAT TTAGCGACCA TAATGGCTAT 150
 TTTACTTAGT AGGATGATC TATATGAGC ATTGCAAA TGTACCGTG GATCAGCAAG 200
 TATTTTAAA CAGGCTAAA TATTTTAAA TTTTATTT TAACTTACCG ATGATGTTC 250
 TTTTACGAGC TTGTAAGTG AATTTTCTA AAAAGGCGT TCAACAGCTC GTATTGTTCG 300
 TTACCAAGAT CCGGGAATAC TTGTAAGCGG TTGCAAAAGA GTGCAAGCGT CTGGCGACCT 350
 TTGTCAGCA CCGGCGCAAG TGTATTTTC ATTTGATGA TGAAGCTTCC ACCGCTCTCG 400
 TTTTACGAGC CCAAGCAAT TTGTAAGCGG TATTTTAAA TTTTATTT TTTTATTT

CGGATGGGTG GAACACTTTC AACCTGACGC TGCAAGGCGA CGTCAAGCGG TTCCGGGGGT 720
 TTGACAACTG GGAAGGCGAT GCGGCTACCG GTTGCGAGGC TTGCTCGAT CAACAACGGC 780
 AATGGATACT CCACATGGCC AAATTGAGCG CTGCGATGGC CAAGCAGGCT CAATATGTCG 840
 CGCAGCTGCA CGTGTGGGCT AGGCGGGAAC ATCCGACTTA TGAAGACATA GTCGGGCTCG 900
 AACGGCTTTA CGCGGAAAAC CCTTCGGCCC GCGACCAAAT TCTCCCGGTG TACGCGGAGT 960
 ATCAGCAGAG GTCGGAGAAG GTGCTGACCG AATACAACAA CAAGGCAGCC CTGGAACCGG 1020
 TAAACCCGCC GAAGCCTCCC CCGGCCATCA AGATCGAGCC GCGCCCGCCT CCGCAAGAGC 1080
 AGGGATTGAT CCTTGGCTTC GTGATGCCGC GTTCTGACCG CTCCGGTGTG ACTCCCGGTA 1140
 CCGGGATGCC AGCCGCAGCG ATGGTTCCGC GTACCGGATC GCGGGGTGGT GGCCTCCCGG 1200
 CTGACACGGC GCGCGAGCTG ACCTCGGCTG GCGCGGAAGC CCGAGCGCTG TCGGGCGAGC 1260
 TCGCGGTCAA AGCGGCATCG CTGGGTGGCG GTGGAGGGCG CCGGGTCCCG TCGGCGCCGT 1320
 TCGGATCCGC GATCGGGGGC GCGAATCCG TCGGGCCCGC TGGCGCTGGT GACATTGCCG 1380
 GCTTAGGCCA GGAAGGGCC GCGCGCCGC CCGCGCTGGG CCGCGGTGGC ATGGGAATGC 1440
 CGATGGGTGC CCGGCATCAG CGACAGGGG CCGCCAAGTC CAAGGGTTCT CAGCAGGAAG 1500
 AGGAGGCGCT CTACCCGAG GATCGGGCAT GCACCGAGG CCGCATTGGT AACCGTCGGC 1560
 CTCAGGACG TAAGGAGTGC AAGTGAGCAT GCACGAATG GACCGGCATG TCGCCCGGGC 1620
 GTTGAAGCTC CCGCGCGGCT TTGAGTGGC GTAGACGGG ACCTCAATC AGATGAACAA 1680
 TGGATGCTTC CCGGCCAGCG ACGAAGCCGA GACCTCGAA GTGACGATCA ATGGGCACCA 1740
 GTGGTTCAGC GCGCTCCGCA TCGAAGATCG TTTCCTGAAG AAGCTGGGTG TCGAGGCGGT 1800
 TGTTAAGCA GTCAAGAGG TGTGTAACAA TGTGAGGCTT GCGGCTGCT GTTATAACGA 1860
 TGGTGTCTT GAGGAGCTGA TGTGTGCTT ATGAGGCTG TCGGCGGCA TGAAGGAAG 1920
 AATGGCTTAA TGTGTTGCTT TGTGTGCTA TATGAGGCA CGGATGAGC GTTCAATGC 1980
 TGTGATGCA TCGCGCCGAG ACGCGCTGA TATGATTTCT CAATGTTTTG ACATGGATCG 2040
 TCGGGCTTTC GAGGCGGCGA TATGCTGCTT TGTGATATT CCGCGAGCTA GTTGTCTTA 2100
 GTTGTGCTA TGTGCTTAA TATGAGCTT GTTGAACA 2160

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:184:

Met Thr Gln Ser Gln Thr Val Thr Val Asp Gln Gln Glu Ile Leu Asn
 1 5 10 15
 Arg Ala Asn Glu Val Glu Ala Pro Met Ala Asp Pro Pro Thr Asp Val
 20 25 30
 Pro Ile Thr Pro Cys Glu Leu Thr Ala Ala Lys Asn Ala Ala Gln Gln
 35 40 45
 Leu Val Leu Ser Ala Asp Asn Met Arg Glu Tyr Leu Ala Ala Gly Ala
 50 55 60
 Lys Glu Arg Gln Arg Leu Ala Thr Ser Leu Arg Asn Ala Ala Lys Ala
 65 70 75 80
 Tyr Gly Glu Val Asp Glu Glu Ala Ala Thr Ala Leu Asp Asn Asp Gly
 85 90 95
 Glu Gly Thr Val Gln Ala Glu Ser Ala Gly Ala Val Gly Gly Asp Ser
 100 105 110
 Ser Ala Glu Leu Thr Asp Thr Pro Arg Val Ala Thr Ala Gly Glu Pro
 115 120 125
 Asn Phe Met Asp Leu Lys Glu Ala Ala Arg Lys Leu Glu Thr Gly Asp
 130 135 140
 Gln Gly Ala Ser Leu Ala His Phe Ala Asp Gly Trp Asn Thr Phe Asn
 145 150 155 160
 Leu Thr Leu Gln Gly Asp Val Lys Arg Phe Arg Gly Phe Asp Asn Trp
 165 170 175
 Gln Gly Asp Ala Ala Thr Ala Trp Glu Ala Ser Leu Asp Gln Gln Arg
 180 185 190
 Lys Thr Gln Leu His Met Ala Lys Leu Ser Ala Ala Met Ala Lys Gln
 195 200 205
 Ala Gln Tyr Val Ala Gln Leu His Val Trp Ala Arg Arg Gln His Pro
 210 215 220
 Thr Tyr Glu Asp Ile Val Gly Leu Gln Arg Leu Tyr Ala Gln Asn Pro
 225 230 235 240

175

Val Asn Pro Pro Lys Pro Pro Pro Ala Ile Lys Ile Asp Pro Pro Pro
 275 280 285

Pro Pro Gln Glu Gln Gly Leu Ile Pro Gly Phe Leu Met Pro Pro Ser
 290 295 300

Asp Gly Ser Gly Val Thr Pro Gly Thr Gly Met Pro Ala Ala Pro Met
 305 310 315 320

Val Pro Pro Thr Gly Ser Pro Gly Gly Gly Leu Pro Ala Asp Thr Ala
 325 330 335

Ala Gln Leu Thr Ser Ala Gly Arg Glu Ala Ala Ala Leu Ser Gly Asp
 340 345 350

Val Ala Val Lys Ala Ala Ser Leu Gly Gly Gly Gly Gly Gly Val
 355 360 365

Pro Ser Ala Pro Leu Gly Ser Ala Ile Gly Gly Ala Glu Ser Val Arg
 370 375 380

Pro Ala Gly Ala Gly Asp Ile Ala Gly Leu Gly Gln Gly Arg Ala Gly
 385 390 395 400

Gly Gly Ala Ala Leu Gly Gly Gly Gly Met Gly Met Pro Met Gly Ala
 405 410 415

Ala His Gln Gly Gln Gly Gly Ala Lys Ser Lys Gly Ser Gln Gln Glu
 420 425 430

Asp Glu Ala Leu Tyr Thr Glu Asp Arg Ala Trp Thr Glu Ala Val Ile
 435 440 445

Gly Asn Arg Arg Arg Gln Asp Ser Lys Glu Ser Lys
 450 455 460

7 INFORMATION FOR SEQ ID NO:185

1. SEQUENCE CHARACTERISTICS
- A. LENGTH: 460 amino acids
 - B. TYPE: amino acid
 - C. STRANDEDNESS:
 - D. TOPOLOGY: linear

81. SEQUENCE DESCRIPTION: SEQ ID NO:185

Ala Gly Asn Val Thr Ser Ala Ser Gly Pro Met Arg Phe Gly Ala Pro
 1 10 15

176

Phe Pro Pro Ser Arg Arg Gln Leu Arg Arg Arg Val Ser Arg Glu Ala
 50 55 60
 Thr Thr Arg Arg Ser Gly Arg Arg Asn His Arg Cys Gly Trp His Pro
 65 70 75 80
 Gly Thr Gly Ser His Thr Gly Ala Val Arg Arg Arg His Gln Glu Ala
 85 90 95
 Arg Asp Gln Ser Leu Leu Leu Arg Arg Arg Gly Arg Val Asp Leu Asp
 100 105 110
 Gly Gly Gly Arg Leu Arg Arg Val Tyr Arg Phe Gln Gly Cys Leu Val
 115 120 125
 Val Val Phe Gly Gln His Leu Leu Arg Pro Leu Leu Ile Leu Arg Val
 130 135 140
 His Arg Glu Asn Leu Val Ala Gly Arg Arg Val Phe Arg Val Lys Pro
 145 150 155 160
 Phe Glu Pro Asp Tyr Val Phe Ile Ser Arg Met Phe Pro Pro Ser Pro
 165 170 175
 His Val Gln Leu Arg Asp Ile Leu Ser Leu Leu Gly His Arg Ser Ala
 180 185 190
 Gln Phe Gly His Val Glu Tyr Pro Leu Pro Leu Leu Ile Glu Arg Ser
 195 200 205
 Leu Ala Ser Gly Ser Arg Ile Ala Phe Pro Val Val Lys Pro Pro Glu
 210 215 220
 Pro Leu Asp Val Ala Leu Gln Arg Gln Val Glu Ser Val Pro Pro Ile
 225 230 235 240
 Arg Lys Val Arg Gln Arg Cys Ala Leu Val Ala Arg Phe Gln Leu Pro
 245 250 255
 Val Arg Phe Phe Gln Ile His Ile Val Ile Phe Thr Gly Arg Ile His
 260 265 270
 His Arg Arg Ile Gly
 275

3. INFORMATION FOR SEQ ID NO:186:

SEQUENCE CHARACTERISTICS:

A. LENGTH: 186 amino acids

177

Arg Val Ala Ala Ser Phe Ile Asp Trp Leu Asp Ser Pro Asp Ser Pro
 1 5 10 15
 Leu Asp Pro Ser Leu Val Ser Ser Leu Leu Asn Ala Val Ser Cys Gly
 20 25 30
 Ala Glu Ser Ser Ala Ser Ser Ser Ala Arg Ser Gly Asn Gly Ser Arg
 35 40 45
 Trp Thr Ser Met Pro Ser Gly Thr Arg Pro Gly Pro Arg Arg Ala Thr
 50 55 60
 Ser Arg Asp Asp Arg Arg Ser Ala Thr Ser Val Ile Pro Ser Arg Arg
 65 70 75 80
 Ser Val Ala Pro Arg Ala Glu Phe Gly Thr Arg Leu Ala Ser His Arg
 85 90 95
 Ala Ser Pro Ser Asn Ala Cys Pro Val Arg Ile Val Thr Ser Ala Ser
 100 105 110
 Gly Arg Pro Ile Ser Ser Pro Pro Ile Val Arg Ser Arg Ser Cys Val
 115 120 125
 Asp Lys Asn Gly Arg Arg Cys Ala Ser Gly Tyr Arg Arg Leu Asn Arg
 130 135 140
 Ala Arg Ser Ser Ser Ile Ala Ala Arg Cys Arg Thr Ile Gly Thr Phe
 145 150 155 160
 Arg Arg Ser Arg Tyr Ser Ala Ser Met Arg Val Ser Thr Asn Ser Pro
 165 170 175
 His Val Thr His Gly Val Ala Pro Gly Val Thr Arg Arg Ile Gly Gly
 180 185 190

INFORMATION FOR SEQ ID NO 187

1. SEQUENCE CHARACTERISTICS

- (A) LENGTH: 190 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO 187

His Val Arg His Gly Met Asp Ile Arg Val Ser Gly Ile Gly His Arg
 1 5

178

35 40 45
 Arg Asn Pro Arg Arg Ser Ser Arg Arg Asp Ala Glu Asp Arg Arg Val
 50 55 60
 Ile Phe Ala Ala Thr Leu Val Ala Val Asp Pro Pro Leu Arg Gly Ala
 65 70 75 80
 Gly Gly Glu Ala Asp Gln Leu Ile Asp Leu Gly Val Cys Arg Arg Gln
 85 90 95
 Ala Gly Arg Val Arg Arg Gly Gln Glu Leu His His Arg His Arg His
 100 105 110
 Gln Gly Ala Ala Pro Asp Leu Arg Arg Arg Arg Arg His Arg Arg Val
 115 120 125
 Gln Gln His Arg Arg Leu Gln Arg Val Arg Gln Leu Arg Arg Tyr Val
 130 135 140
 Gln Thr Ala His His Arg Arg Phe Ala Arg Thr Asp Arg Val Arg His
 145 150 155 160
 His Val Arg Gly Pro Ser Asn His Arg Arg Arg Arg Val Tyr Arg Gly
 165 170 175
 Arg His Ser Gly Ala Gly Gly Cys Pro Ala Gly Gly Ala Gly Ser Val
 180 185 190
 Gly Gly Ser Ala
 195

2. INFORMATION FOR SEQ ID NO: 134

2.1. SEQUENCE CHARACTERISTICS

- (A) LENGTH: 211 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

2.2. SEQUENCE DESCRIPTION: SEQ ID NO: 134

Val Arg Cys Gly Thr Leu Val Ile Val Ile Met Val Glu Phe Leu Thr
 5 10 15
 Ser Thr Asn Ala Pro Ser Leu Thr Ser Ala Tyr Ala Glu Val Asp Lys
 20 25 30
 Leu Ile Val Leu Ile Val Arg Thr Ala Lys Arg Thr Ile Asn Gly Thr
 35 40

65 70 75 80

Ala Glu Tyr Arg Asp Arg Arg Lys Val Pro Ile Val Arg Gln Arg Ala
85 90 95

Ala Ile Glu Glu Leu Arg Ala Arg Phe Asn Leu Arg Tyr Pro Leu Ala
100 105 110

His Leu Arg Pro Phe Leu Ser Thr His Glu Arg Asp Leu Thr Met Gly
115 120 125

Gly Glu Glu Ile Gly Leu Pro Asp Ala Glu Val Thr Ile Arg Thr Gly
130 135 140

Gln Ala Leu Leu Gly Asp Ala Arg Trp Leu Ala Ser Leu Val Pro Asn
145 150 155 160

Ser Ala Arg Gly Ala Thr Leu Arg Arg Leu Gly Ile Thr Asp Val Ala
165 170 175

Asp Leu Arg Ser Ser Arg Glu Val Ala Arg Arg Gly Pro Gly Arg Val
180 185 190

Pro Asp Gly Ile Asp Val His Leu Leu Pro Phe Pro Asp Leu Ala Asp
195 200 205

Asp Asp Ala Asp Asp Ser Ala Pro His Glu Thr Ala Phe Lys Arg Leu
210 215 220

Leu Thr Asn Asp Gly Ser Asn Gly Glu Ser Gly Glu Ser Ser Gln Ser
225 230 235 240

Ile Asn Asp Ala Ala Thr Arg Tyr Met Thr Asp Glu Tyr Arg Gln Phe
245 250 255

Pro Thr Arg Asn Gly Ala Gln Arg Ala Leu His Arg Val Val Thr Leu
260 265 270

Leu Ala Ala Ile Arg Pro Val Leu Thr His Tyr Pro Ala Gly Lys Asp
275 280 285

Arg Thr Gly Pro Val Val Ala Leu Val Leu Glu Ala Val Gly Leu Asn
290 295 300

Arg Asp Val Ile Val Ala Asp
305 310

2 INFORMATION FOR SEC ID NO 184

1. *Phragmites australis* (Cav.) Trin. ex Steud.

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:189:

CTCGTGCCGA TTCCGCACGA GCTGAGCAGC CCAAGGGGCC GTTCGGCGAA GTCATCGAGG	60
CATTGCCCCA CGGGCTGGCC GGCAAGGGTA AGCAAATCAA CACCACGCTG AACAGCCTGT	120
CGCAGGCGTT GAACGCCTTG AATGAGGGCC GCGGCGACTT CTTCCGGGTG GTACGCAGCC	180
TGGCGCTATT CGTCAACGCG CTACATCAGG ACGACCAACA GTTCGTCCGG TTGAACAAGA	240
ACCTTGCCGA GTTCACCGAC AGGTTGACCC ACTCCGATGC GGACCTGTGG AACGCCATCC	300
AGCAATTCCA CAGCTTGCTC GCGGTCCGCG GCGGCTTCTT CGCCAAGAAC CGCGAGGTGC	360
TGACGCATGA CGTCAATAAT CTCGCGACCG TGACCACCAC GTTGCTGCAG CCGGATCCGT	420
TGGATCCCTT CGAGACGCTC TTGACATCTT TCCCGACGCT GGCGGCGAAC ATTAACCAGC	480
TTTACCATCC GACACACGGT GCGGTGGTGT CCGTTTCCGC GTTCACGAAT TTGCCAACC	540
CGATGAGTT CATCTGCAGC TCGATTCAGG CGCGTAGCTG GCTCGGTTAT CAAGAGTCGG	600
CGGAATCTG TCCCACTAT CTGGCGCCAG TCCTCGATGC GATCAAGTTC AACTACTTTC	660
CGTTGGGCTT GAACGTGGCC AGCAGCGGCT CGACACTGCC TAAAGAGATC GCGTACTCCG	720
AGCGCGGCTT CGAGCGCGCC AACGGGTACA AGGACACCAC GGTGCGCGGC ATCTGGGTGC	780
CGGATACGCT GTTCTACAC CGCAACACCG AGCGCGGTTG GGTGGTGGCA CCGGGGATGC	840
AGGGGCTTCA GGTGGGACCG ATCAGCGAGG GTTGTGTGAC CGCGGAGTCC CTGGCGGAAC	900
TCATGGGTGG TCGGATATC CGCGCTCGCT CCGTAGGGGT GGAACCGCG CCGGACCGCT	960
CGAATGCCTA CGACGAGTAC CGCGTCTGC CGCGGATCG TTTACAGGCC CCACAGGTGC	1020
CGATACACCG CGCGCTCGCT GCGCGCGAGC TAATCGCGCG TCGGTGCCA CCGGTCTTGG	1080
CGCGCATGCT GTTCCGAAGA GATCGCGCG GAGGCTTGA AACTTGGAC TACATGGGCC	1140
CTTCTTCTCT CTGCGCGCG GTGCGACCT TCTGTCTCTT TCTCTCTCT AGCGCGCGCG	1200
CTGGAAGGAT AGGCGATCG CAGCTTTTGA TACGGCGAT TACGCGCTG GCGTTGATCG	1260
CTGATCTCT CCGCATCTCG TGGTACCGCA CAGAACATCG TCTCATAGAC ATCGCGCTTGT	1320
TCGAGAACCG AGCGGTGCG CAGGCGGACA TGACGATGAT GTTGTCTTGC CTCGGGCTGT	1380
CTGCTCTCTT CTGCTCTCTT TGAAGTACCT TCGAGCAAT GTTCTCTCTT TCGCGATCT	1440
AATCTCTCTT CCGATCTCTT TCGCTCTCTT TCGCTCTCTT TCGCTCTCTT TCGCTCTCTT	1500

TGGCGACCCG GCTGGCAATC ATGGGUATCG GCATGGGCTG CTCCATGATG CCACTGTCCG 1680
 GGGCGGCAGT GCAGACCCCTG GGGCCACATC AGATCGCTCG CGGTTCGACG CTGATCAGCG 1740
 TCAACCAGCA GGTGGGCGGT TCGATAGGGA CCGCACTGAT GTCGGTGCTG CTCACCTACC 1800
 AGTTCAATCA CAGCGAAATC ATCGCTACTG CAAAGAAACT CGCACTGACC CCAGAGAGTG 1860
 GCGCCGGGGCG GGGGGCGGCG GTTGACCCCTT CCTCGCTACC GCGCCAAACC AACTTCGCGG 1920
 CCCAACTGCT GCATGACCTT TCGCACGCTT ACCCGGTGGT ATTCTGTATA GCGACCGCGC 1980
 TAGTGGTCTC GACGCTGATC CCGCGCGCAT TCCTGCCGAA ACAGCAGGCT AGTCATCGAA 2040
 GAGCACCGTT GGTATCCCA TACCTCTCTC TT 2072

2. INFORMATION FOR SEQ ID NO:190:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1923 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:190:

TCACCCCGCA GAAGTCGTTG GTGACGAGC TGGACATCGA CTGGCTGTCTG ATGGTCGAGA 60
 TGGCGGTGCA GACCGAGGAG AACTAGCGCG TCAAGATCCC TGACGAGGAG CTCGCGCGTC 120
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 180
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 240
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 300
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 360
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 420
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 480
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 540
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 600
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 660
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 720

INFORMATION FOR SEC ID NO. 141

1. GENERAL CHARACTERISTICS

- A LENGTH: 1095 base pairs
B TYPE: nucleic acid
C STRANDEDNESS: single
D TOPOLOGY: linear

01 SEQUENCE DESCRIPTION: SEQ ID NO: 191

TO: DIRECTOR, FBI (100-442654) FROM: SAC, NEW YORK (100-100000) (P)
SUBJECT: MURDER OF MARTIN LUTHER KING, JR.;
RE: NEW YORK TELETYPE TO BUREAU, APRIL 11, 1968.

GGGGATGGTT CAGACGTAAC GGTTGGCTAG GTCGAAACCC GCGCCAGGGC CGCTGGACGG 300
 GCTCATGGCA GCGAAATTAG AAAACCCGGG ATATTGTCCG CGGATTGTCA TACGATGCTG 360
 AGTGCTTGGT GGTTCGTGTT TAGCCATTGA GTGTGGATGT GTTGAGACCC TGGCCTGGAA 420
 GGGGACAACG TGCTTTTGCC TCTTGGTCCG CCTTTGCCCG CCGACGCGGT GGTGGCGAAA 480
 CCGGCTGAGT CCGGAATGCT CCGCGGGTTG TCGGTTCCGC TCAGCTGGGG AGTGGCTGTG 540
 CCACCCGATG ATTATGACCA CTGGGCGCCT GCGCCGGAGG ACGGCGCCGA TGTCGATGTC 600
 CAGGCGGCCC AAGGGGCGGA GCGAGAGGCC GCGGCCATGG ACGAGTGGGA TGAGTGGCAG 660
 GCGTGAACG AGTGGGTGGC GGAGAACCCT GAACCTGCTT TTGAGGTGGC ACGGAGTAGC 720
 AGCAGCGTGA TTCCGCTTTC TCCGCGCGCC GGCTAGGAGA GGGGGCGCAG ACTGTCGTTA 780
 TTGACCACT GATCGGCGGT CTCGGTGTTC CCGCGGCCGG CTATGACAAAC AGTCAATGTG 840
 CATGACAAGT TACAGGTATT AGGTCCAGGT TCAACAAGGA GACAGGCAAC ATGGCAACAC 900
 GTTTTATGAC GGATCCGCAC GCGATGCGGG ACATGGCGGG CCGTTTTGAG GTGCACGCCC 960
 AGACGGTGGG GGACGAGGCT GCGCGGATGT GGGCGTCCGC GCAAAACATC TCCGCGCGGG 1020
 GCTGAGTGG CATGGCCGAG GCGACCTGCG TAGAC 1055

12. INFORMATION FOR SEQ ID NO:192:

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 150 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

11. SEQUENCE DESCRIPTION: SEQ ID NO:192

TGGCTGCTTT GTTGGGATAA TGGGCGCGG TGGGCTGGA TGGCTGCGG TGGGCTGCTG 97
 TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG 157
 TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG 187
 TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG 247
 TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG 307
 TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG TGGGCTGCTG 367

184

- (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:193:

```

AACGGGCCCC TGGCCACCGC TCCTCTAAGG GCTCTCGTTG GTCGCATGAA GTGCTGGAAG      60
GATGCATCTT GGCAGATTCC CGCCAGAGCA AAACAGCCGC TAGTCCTAGT CCGAGTCGCC      120
CGCAAAGTTC CTCGAATAAC TCCGTACCCG GAGCGCCAAA CCGGGTCTCC TTCGCTAAGC      180
TGCGCGAACC ACTTGAGGTT CCGGGACTCC TTGACGTCCA GACCGATTCC TTCGAGTGGC      240
TGATCGGTTC GCGCGGCTGG CCGGAATCCG CCGCCGAGCG GGGTGATGTC AATCCAGTGG      300
CTGCGCTGGA AGAGGTGCTC TACGAGCTGT CTCGATCGA CGACTTCTCC      360
  
```

(2) INFORMATION FOR SEQ ID NO:194:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 679 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:194:

```

Glu Gln Pro Lys Gly Pro Phe Gly Glu Val Ile Gln Ala Phe Ala Asp      1
1      5      10      15
Gly Leu Ala Gly Lys Gly Lys Gln Ile Asn Thr Thr Leu Asn Ser Leu      20
20      25      30
Ser Gln Ala Leu Asn Ala Leu Asn Glu Gly Arg Gly Asp Phe Phe Ala      35
35      40      45
Val Val Arg Ser Leu Ala Leu Phe Val Asn Ala Leu His Gln Asp Asp      50
50      55      60
Gln Gln Phe Val Ala Leu Asn Lys Asn Leu Ala Gln Phe Thr Asp Arg      65
65      70      75
Leu Thr His Ser Asp Ala Asp Leu Ser Asn Ala Ile Gln Gln Phe Asp      80
80      85      90
Ser Leu Leu Ala Val Ala Arg Pro Phe Phe Ala Lys Asn Arg Gln Val      95
95      100      105      110
Leu Thr His Asp Val Asn Asn Leu Ala Thr Val Thr Thr -      115
115
  
```

185

Thr Leu Ala Ala Asn Ile Asn Gln Leu Tyr His Pro Thr His Gly Gly
 145 150 155 160
 Val Val Ser Leu Ser Ala Phe Thr Asn Phe Ala Asn Pro Met Glu Phe
 165 170 175
 Ile Cys Ser Ser Ile Gln Ala Gly Ser Arg Leu Gly Tyr Gln Glu Ser
 180 185 190
 Ala Glu Leu Cys Ala Gln Tyr Leu Ala Pro Val Leu Asp Ala Ile Lys
 195 200 205
 Phe Asn Tyr Phe Pro Phe Gly Leu Asn Val Ala Ser Thr Ala Ser Thr
 210 215 220
 Leu Pro Lys Glu Ile Ala Tyr Ser Glu Pro Arg Leu Gln Pro Pro Asn
 225 230 235 240
 Gly Tyr Lys Asp Thr Thr Val Pro Gly Ile Trp Val Pro Asp Thr Pro
 245 250 255
 Leu Ser His Arg Asn Thr Gln Pro Gly Trp Val Val Ala Pro Gly Met
 260 265 270
 Gln Gly Val Gln Val Gly Pro Ile Thr Gln Gly Leu Leu Thr Pro Glu
 275 280 285
 Ser Leu Ala Glu Leu Met Gly Gly Pro Asp Ile Ala Pro Pro Ser Ser
 290 295 300
 Gly Leu Gln Thr Pro Pro Gly Pro Pro Asn Ala Tyr Asp Glu Tyr Pro
 305 310 315 320
 Val Leu Pro Pro Ile Gly Leu Gln Ala Pro Gln Val Pro Ile Pro Pro
 325 330 335
 Pro Pro Pro Gly Pro Asp Val Ile Pro Gly Pro Val Pro Pro Val Leu
 340 345 350
 Ala Ala Ile Val Phe Pro Arg Asp Arg Pro Ala Ala Ser Glu Asn Phe
 355 360 365
 Asp Tyr Met Gly Leu Leu Leu Leu Ser Pro Gly Leu Ala Thr Phe Leu
 370 375 380
 Phe Gly Val Ser Ser Ser Pro Ala Arg Gly Thr Met Ala Asp Arg His
 385 390 395 400
 Val Leu Ile Pro Ala Ile Thr Gly Leu Ala Leu Ile Ala Ala Phe Val
 405

435 440 445

Ser Leu Gly Leu Phe Gly Ser Phe Leu Leu Leu Pro Ser Tyr Leu Gln
450 455 460

Gln Val Leu His Gln Ser Pro Met Gln Ser Gly Val His Ile Ile Pro
465 470 475 480

Gln Gly Leu Gly Ala Met Leu Ala Met Pro Ile Ala Gly Ala Met Met
485 490 495

Asp Arg Arg Gly Pro Ala Lys Ile Val Leu Val Gly Ile Met Leu Ile
500 505 510

Ala Ala Gly Leu Gly Thr Phe Ala Phe Gly Val Ala Arg Gln Ala Asp
515 520 525

Tyr Leu Pro Ile Leu Pro Thr Gly Leu Ala Ile Met Gly Met Gly Met
530 535 540

Gly Cys Ser Met Met Pro Leu Ser Gly Ala Ala Val Gln Thr Leu Ala
545 550 555 560

Pro His Gln Ile Ala Arg Gly Ser Thr Leu Ile Ser Val Asn Gln Gln
565 570 575

Val Gly Gly Ser Ile Gly Thr Ala Leu Met Ser Val Leu Leu Thr Tyr
580 585 590

Gln Phe Asn His Ser Gln Ile Ile Ala Thr Ala Lys Lys Val Ala Leu
595 600 605

Thr Pro Gln Ser Gly Ala Gly Arg Gly Ala Ala Val Asp Pro Ser Ser
610 615 620

Leu Pro Arg Gln Thr Asn Phe Ala Ala Gln Leu Leu His Asp Leu Ser
625 630 635 640

His Val Tyr Ala Val Val Phe Val Ile Ala Thr Ala Leu Val Val Ser
645 650 655

Thr Leu Ile Gln Ala Ala Pro Leu Arg Leu Gln Gln Ala Ser His Arg
660 665 670

Arg Ala Pro Leu Leu Ser Ala
675

2 INFORMATION FOR SEQ ID NO 188

RESEARCH DESIGN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:195:

Thr Pro Glu Lys Ser Phe Val Asp Asp Leu Asp Ile Asp Ser Leu Ser
 1 5 10 15
 Met Val Glu Ile Ala Val Gln Thr Glu Asp Lys Tyr Gly Val Lys Ile
 20 25 30
 Pro Asp Glu Asp Leu Ala Gly Leu Arg Thr Val Gly Asp Val Val Ala
 35 40 45
 Tyr Ile Gln Lys Leu Glu Glu Glu Asn Pro Glu Ala Ala Gln Ala Leu
 50 55 60
 Arg Ala Lys Ile Glu Ser Glu Asn Pro Asp Ala Ala Arg Ala Asp Arg
 65 70 75 80
 Cys Val Ser Pro Thr Ser Gln Ala Arg Asp Ala Arg Arg Pro Leu Ala
 85 90 95
 Arg Ser Ala Arg Leu Ala Cys Arg Arg Leu Pro Ala Ser Val Pro Thr
 100 105 110
 Thr Arg Arg Asp Pro Arg Glu Arg
 115 120

B. INFORMATION FOR SEQ ID NO:196:

- (A) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 99 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

xi. SEQUENCE DESCRIPTION: SEQ ID NO:196:

Leu Ala Cys Gln Cys His Arg Arg Tyr Asp Val Gly Ile Gln Phe Arg
 1 5 10 15
 His Pro Ala Gly Pro Val Ala Thr Gln Ser Gly Pro Pro Gly Ser Ser
 20 25 30
 His Ala His His Arg Gln Val Arg Ala Gln Lys Gly Ala Gly Phe Leu
 35 40 45
 His Arg Arg Pro Ala Val Ser Gly Ala Leu Pro Pro Asn Asn Ala Ser
 50 55 60
 Pro Gly His Arg Ser Arg Ala Ala Asn Ser Gln Arg His Leu Leu Ala
 65 70 75 80 85 90 95

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 119 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:197:

Ala Ser Leu Leu Ala Tyr Ser Ala Ala Ala Ser Thr Ala Leu Ala
 1 5 10 15
 Val Ala Cys Val Arg Ala Asp His Arg Asp Arg Arg Thr Ile Arg Asp
 20 25 30
 His Leu Ala Met Ile His Leu Ala Gln Leu Val Thr Gln Pro Pro Gly
 35 40 45
 Gly Val Arg Gln Arg Leu His His Leu Gly Ile Ala Val Ala Pro Gln
 50 55 60
 Pro Gln Glu Val Val Val Leu Ala His His Leu Val Thr Gly Thr Gly
 65 70 75 80
 Glu Val Gln Gly Glu Gly Arg His Val Ala Ala Glu Val Val Asp Pro
 85 90 95
 Glu Asn Gln Ile Leu Arg Gln Val Leu Gly Pro Ala Pro His Asp Lys
 100 105 110
 Pro Asp Ala Gly Ile Gly Gln
 115

2 INFORMATION FOR SEQ ID NO:198:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 116 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:198:

Arg Ala Arg Gly His Arg Ser Ser Lys Gly Ser Arg Trp Ser His Glu
 1 5 10 15
 Val Leu Glu Gly Cys Ile Leu Ala Asp Ser Arg Gln Ser Lys Thr Ala
 20 25 30
 Ala Ser Pro Ser Pro ...
 35

Glu Val Pro Gly Leu Leu Asp Val Gln Thr Asp Ser Phe Glu Trp Leu
 65 70 75 80
 Ile Gly Ser Pro Arg Trp Arg Glu Ser Ala Ala Glu Arg Gly Asp Val
 85 90 95
 Asn Pro Val Gly Gly Leu Glu Glu Val Leu Tyr Glu Leu Ser Pro Ile
 100 105 110
 Glu Asp Phe Ser
 115

(2) INFORMATION FOR SEQ ID NO:199:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 811 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO:199:

TGCTACGCAG CAATCGCTTT GGTGACAGAT GTGGATGCCG GCCTCGCTGC TGGCGATGGC 60
 GTGAAAGCCG CCGAGCTGTT CCGCCGATTC GGGGAGAACA TCGAACTGCT CAAAAGGCTG 120
 GTGCGGGCCG CCATCGATCG GGTGCGCGAC GAGGCGACGT GCACGCACTG TCAACACCAC 180
 GCGCGTGTTC GGTTCGCTTT CGAGCTGCCA TGAGGCTGCT GGTGACCGGC GCGGCGGGCT 240
 TCATCGGGTC GCGCGTGGAT GCGGCGTTAC GGGCTGCGGG TCACGACGTG GTGGGCGTGC 300
 ACCCGCTGCT GCGGCGCGCG CACGCGCCAA ACCCGCTGCT TCGACCGCGG TCGCAGCGGG 360
 TCGAGCTGCG CGACTCCAGC GCGCTGGCCC GGTGCTTGGC CGGTGTCCAT CTGGTGTGTC 420
 ACCAGGCGCG CATGGTGGGT GCGGCGGTCA ACCCGCGCGA CGUACCGCGC TATGGCGGCG 480
 ACGAGGATTT CCGCAGCAGG GTGCTGCTCG CCGAGATGTT CGGCGCGCGG GTGCGCGCTT 540
 TGGTGTGCGG GTGCTGAGG GTGCTTACG CCGAGGCGCG GTATGCTGT CCGCAGCAGG 600
 GAGGCTGCGA CCGCTGCGG GGTGCGCGAG CCGAGCTGCA GTATGCTGT TCGCAGCAGG 660
 GTTTCGCGGG GTGCGCGGAG CCGCTGCTGT CCGAGTGTGT GTATGAGAT GCGGCGCTGC 720
 GTGCGCGGAG CCGCTGACCG CCGCAGAGAG CCGCAGGAG GTATGCGCG TCGGCTGCTG 780
 GGAAGCAAT GCGGCTGCGG GTATGCTGT 840

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:200:

```

GTCCCGCGAT GTGGCCGAGC ATGACTTTTCG SCAACACCCG CGTAGTAGTC GAAGATATCG      60
GACTTTGTTCG TCCCGGTGGC GGGATAGAGC ACC1GTCCGC GTTGGTCAGC GTCACCCGTT      120
GCTCGGACGC CGAACCCATG CTTTCAACGT AGCCTGTCCG TCACACAAGT CGCGAGCGTA      180
ACGTCACGGT CAAATATCGC GTGGAATTTG GCCGTGACGT TCCGCTCCCG GACAATCAAG      240
GCATACTCAC TTACATCCGA GCCATTTGGA CCGGTTTCGAT CGCCTTCGGG CTGGTGAACG      300
TCCCGGTCAA GGTGTACAGC GGTACCGAG ACCACGACAT CAGGTTCAC CAGGTGCACG      360
CCAAGGACAA CGGACCCATC CGGTACAAGC GCCTCTCCGA GCGCTCTGGC GAGGTGCTCG      420
ACTACCGCGA TCTTGCCCGG GCCTACGAGT CCGCGGACCG CCAAATGGTG GCGATCACCG      480
ACGACGACAT CGCCAGCTTG CCTGAAGAAG GCAGCCCGGA GATCGAGGTG TTGGAGTTCC      540
TCCCGCCCGC CGACGTGGAC CCGATGATGT TCGACCCGAG CTACTTTTTG GAGCCTGATT      600
CGAAGTCGTC GAAATCGTAT GTGCTGCTGG GTAAGACACT CGCCGAGACC GACCGGATGG      660
CGATCGTGGA TCGCCCGACC GCGCGTGAAT GCAGGAAAAA TAAGAGCCCG TATCCACAAT      720
TCCCGCTCGA GCTCGGCTAC CACCAACGGT AGAAGGATCG AGACATTCGG GAGCTGAAGT      780
TCCCGCTAT AGAAGCTGCT TCCCGGATT ATCAAACGCA AATACGCTT ACTCATGCGA      840
TCCCGCTCG TCCCGGATG CGACCTTTT TCCAGCTCG AUGGCTTCC CCGCGACCTC      900
AAGTGGGAT GCATCCGAGC GTTCCCGGA ACCCGGTTCC GCGGGGTCCG CTCATCGCTT      960
ATCTCT

```

INFORMATION FOR SEQ ID NO 201

SEQUENCE CHARACTERISTICS:
A. LENGTH: 2347 base pairs
B. TYPE: nucleic acid
C. STRANDEDNESS: single
D. TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:201

CGCATCCGCA GCGAATA 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819 1820 1821 1822 1823 1824 1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840 1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2

CTCGCGCGCT GCGACCGTTG CCGCGCTTGC CGATCAACAT GCGGCTGGCG CCACCGTTGC 240
 CACCCACGCG ACCGGCTCCG CCGACCGCGC CGACACCAAG CGAGCTGCCG CCGGAGCCAC 300
 CATCACCACC TACGCCACCG ACCGCGCAGA CACCAGCGAC CGGGTCTTCG TGAAACGTCC 360
 CGGTGCCACC ACCGCGCGCG TTACCGCCAA CCCCACCGGC AACGCGGGCG CCGCCATCCC 420
 CGCGGGCCCC GCGGTGCCG CCGTTGCCG CGTTGCCGAA CAACAACCG CCGGCGCGCG 480
 CGTTGCCGCG CCGCGCGCG GTCCCGCGCG CCGCGCGAC GCCAAGGCG CTGCCGCGCT 540
 TGCGCGCATC ACCACCGTTG CCGCGACCA CATCGGGTTC TGCCTCGGG TCTGGGCTGT 600
 CAAACCTCGG GATGCCAGCG TTGCCCGCGC TTCCCGCGG CCCCCCGTG GCGCGCTAC 660
 CACCGATACC ACCGCGCGCA CCGCGCGCAC CGTTGCCGCG ATCACCGAAT AGCAACCGCG 720
 CGGCGCGACG ATTGCCGCGA GTCGCGCTG CGCGACCGTC GCGCGCGGAG GCGGCACTGG 780
 CAGCGCGCTT ACCACGAAA CCGCGCTAC CACCGGTAGA GGTGGCACTG GCGATGTGA 840
 CGAAAGCGCG GCGTCCGGCG CCGCGCTAC CACCGCGACT GCGGGCGGCT ACACCGTCGG •••900
 ACCCGTTGCC ACCATCACC CGAAAGCGCG TCGCAATGTC GCGGTGCGCG ACTCCGCGCT 960
 CCGCGCGCTT GCGCGCGCG CGACCGCGAG CGCGGTAGC GCGGTACCA CCGGCACCGC 1020
 CGGTGCCCTT GCGCGAGCGT GCGGTGCGCG TGGCACCGTC GCGCGCGGTG CCACCGGTCC 1080
 CGGTGCCGCG AGTCCCATGG CCGCGCGTGC CCGGTGCGCG GCGGTGTTGA TCACCGATGC 1140
 CGGACACATC TCGCGGGCTG TCGCGGTGC TCGCGCGCGG GCGGGGCTG CGATTGACCG 1200
 CGTTTGCGCG GCGGAGGCG GCGCGCGCG TACCACCGCG CCGCGCATGG CGGAACAGCG 1260
 TCGGTGCGCG GCGGTACCG CCGCGACCG CGATGCGTGC GCGACGCTG GTGCGCGCGA 1320
 TCGGTGCGCG GCGCGCTG CCGCGACCG AGCGCGCTT CGACCGCGA CGCGCGCGCG 1380
 TCGGTGCGCG GCGCGCGCG GCGGTGCGCG TCGCGCAT TACCGCGCG GCGGTGCGCG 1440
 TCGGTGCGCG TCGACCGAC CGTCCAGCG TCGGTGCGCG GCGGTGCGCG AACAGCAAGC 1500
 TCGGTGCGCG GCGAGGCTGC CCGGTGCGCG TCGGTGCGCG GCGGTGCGCG ATCAACGCG 1560
 TCGGTGCGCG GCGGTGCGCG GCGGTGCGCG TCGGTGCGCG GCGGTGCGCG ATCAACGCG 1620
 TCGGTGCGCG GCGGTGCGCG GCGGTGCGCG TCGGTGCGCG GCGGTGCGCG ATCAACGCG 1680

CCAGCAGCTG CGGCGTCGCG ATCAGCAAGG ACACCTCGCA CCTCCGGATA CCCCATATCG 1920
 CCGCACCGTG TCCCCAGCGG CCACGTGACC TTTGGTCGCT GGCTGGCGGC CCTGACTATG 1980
 GCGCGGACGG CCTTCGTTCT GATTGCCCCC GCGCGCGCAGC TTGTTGCGCG AGTTGAAGAC 2040
 GGGAGGACAG GCGGASCTTG GTGTAGACGT GGGTCAAGTG GGAATGCACG GTCCGCGGCG 2100
 AGATGAATAG GCGGACGCGG ATCTCCTTGT TGCTGAGTCC CTCACCGACC AGTAGAGCCA 2160
 CCTCAAGCTC TGTCGGTGTC AACGCGCCCC AGCCACTTGT CGGGCGTTTC CGTGACCGGC 2220
 GCGCTCGTTG CCGGTACCGG ATCGCCTCAT CGATCGATAA CGCAGTTCCT TCGGCCCAGG 2280
 CATCCTCGAA CTCGCTGTCA CCGATGGATT TTCGAAGGCT GGCTAGCGAC GAGTTACAGC 2340
 CCGCCTGGTA GATCCCGAAG CGGACCG 2367

(2) INFORMATION FOR SEQ ID NO:202:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 376 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:202:

Gln Pro Ala Gly Ala Thr Ile Ala Ala Ser Ser Pro Cys Ala Thr Val
 1 5 10 15
 Gly Ala Gly Gly Gly Thr Gly Ser Pro Val Thr Thr Glu Thr Ala Ala
 20 25 30
 Thr Thr Gly Arg Gly Gly Ser Gly Asp Val Tyr Glu Ser Ala Ala Ser
 35 40 45
 Gly Ala Ala Ala Thr Thr Ser Thr Ala Gly Gly Tyr Thr Val Gly Pro
 50 55 60
 Val Ala Thr Ile Thr Ala Lys Gly Ala Arg Asn Val Ala Leu Arg Asp
 65 70 75 80
 Ser Ala Val Ala Ala Val Ala Ala Ala Thr Gly Ser Gly Gly Thr
 85 90 95
 Ala Val Thr Thr Gly Thr Ala Gly Ser Leu Ala Ala Ala Cys Arg Arg
 100 105 110

His Ile Cys Arg Ala Val Pro Gly Ala Gly Arg Gly Ala Gly Arg Gly
 145 150 155 160
 Ile Asp Pro Val Cys Pro Gly Glu Ala Gly Ala Ala Gly Thr Thr Gly
 165 170 175
 Ala Ala Met Ala Glu Gln Pro Gly Val Ala Ala Val Thr Ala Arg Thr
 180 185 190
 Pro Asp Ala Cys Gly His Ala Gly Ala Ala Asp Thr Ala Val Ala Ala
 195 200 205
 Val Ala Pro Gln Pro Pro Pro Val Pro Thr Gly Thr Ala Gly Arg Ala
 210 215 220
 Gly Thr Thr Gly Pro Ala Val Ala Ala Val Ala Asp Gln Pro Gly Arg
 225 230 235 240
 Ala Ser Ala Ala Ala Gly Leu Thr Glu Pro Ala Ser Arg Ala Val Ala
 245 250 255
 Thr Val Ala Lys Gln Gln Pro Ala Gly Arg Ala Arg Leu Pro Gly Cys
 260 265 270
 Arg Pro Val Gly Ala Val Ser Asp Gln Arg Ala Pro Gln Lys Arg Leu
 275 280 285
 Gly Gly Arg Ile His Arg Thr Gln Gln Thr Pro Leu Asn Ser Gly Phe
 290 295 300
 Ser Ala Gly Ile Pro Thr Arg Gly Arg Ser Gln Arg Leu His Lys Leu
 305 310 315 320
 Leu Val Lys Arg Cys His Leu Tyr Ala Gln Arg Leu Ile Leu Pro Ser
 325 330 335
 Met Gly Pro Glu Gln Pro Arg Asn Arg Arg Arg His Phe Ile Gly Ser
 340 345 350
 Arg Ser His His Phe Arg Arg Ala Arg Arg Arg Ile Arg Ile Ser Arg
 355 360 365
 Ala His Leu Arg Thr Asn Ser Arg
 370 375

2 INFORMATION FOR SEQ ID NO:203

1 SEQUENCE CHARACTERISTICS

1.1 LENGTH: 375 base pairs

GGCCAAAACG CCGCGGCGAT CGCGGCCACC GAGGCGCGCT ACGACCAGAT GTGGGCCACG 60
GACGTGGCGG CGATGTTTGG CTACCATGCC GGGGCTTCGG CGGCCGTCTC GGCCTTGACA 120
CCGTTCGGCC AGGCGCTGCC GACCGTGGCG GCGGGCGGTG CGCTGGTCAG CGCGGCCGCG 180
GCTCAGGTGA CCACGCGGGT CTTCGCAAC CTGGGCTTGG CGAACGTCCG CGAGGGCAAC 240
GTCCGCAACG GTAATGTCCG GAACTTCAAT CTCGGCTCGG CCAACATCGG CAACGGCAAC 300
ATCGGCAGCG GCAACATCGG CAGCTCCAAC ATCGGGTTTG GCAACGTGGG TCCTGGGTTG 360
ACCGCAGCGC TGAACAACAT CGGTTTCGGC AACACCGGCA GCAACAACAT CGGGTTTGGC 420
AACACCGGCA GCAACAACAT CGGGTTCGGC AATACCGGAC ACCGCAACCG AGGTATCGGG 480
CTCACGGSTA GCGGTTTGGT GGGGTTCCGC GGCCTGAACG CGGGCACCGG CAACATCGGT 540
GTGTTCAACT CGGGCACCGG AAACGTCCGC ATCGGCAACT CGGGTACCGG GAACTGGGGC 600
ATTGGCAACT CGGGCAACAG CTACAACACC GGTTTTGGCA ACTCCGCGCA CGCCAACACG 660
GGCTTCTTCA ACTCCGGAAT AGCCAACACC GCGGTCCGCA ACGCCGGCAA CTACAACACC 720
GGTAGCTACA ACCCGGGCAA CAGCAATACC GCGGGCTTCA ACATGGGGCA GTACAACACG 780
GGGTACCTGA ACAGCGGCAA CTACAACACC GGTTCGGCAA ACTCCGGCAA TGTCAACACC 840
GGGCGCTTCA TTAAGGCAA GTTCAACAAC GGTCTCTTGT GCGCGGGCGA CCACCAAGGC 900
GTGATTTTGG GAGGCGCGCG GTTCTTCAAC TCGACCACTG CGCGCTCGTG GGGATTCTTC 960
AAGAGCGGTG TCGGTACCGC CTCGCGTTTC TGAACCTCG GTGCAACAA TTCTCGCTTC 1020
TTCAACTCTT GTGCGGGGGC CATCGGTAAC TCGGGCTCG CAAACGCGCG CGTGCTGGTA 1080
TTCGGCTGTA TCAACTCGCG CAACACCGTA TCGGTTTGT TCAACATGAG CCGGTGCGGT 1140
ATCAACACCG TCGGTTGAT CTCGCGTTTC TTAACACCG TAAGCAACAT CTCGGGATTT 1200
TTGCTGCGCG TACGCTTGT CATCTCGCG TTAACAACT TCGGCTCGT GAACATTCTC 1260
GCGACCGCA ATATCGGCA TTAACACAT TCGGCGCGG GAAACGTCCG TGAATTCAAC 1320
ATGCTTGGCA GCGGCAACGT CGGCACGCA AACATCTTGG GCAGCGGCAA CGTCGCGAGT 1380
TTCAACATCG GCACTGAAA CATCGGATA TTAATCTTC TTGCGGAA GCTGGGAAA 1440
TAAACATCG CATCGGAAA CTGCGGATG TAACATCTC TTTCTGAAA CCGTGGGAA

AACCAGCAGG GCTTCAATAT TGCTAGCGGC TGGAACTCGG GCACCGGCAG CAGCGGCCTG 1680
 TTCAATTCGG GCACCAATAA GGTGGGCATC TTCAACGCGG GCACCGGAAA CGTCGGCATC 1740
 GCAAACTCGG GCACCGGGAA CTGGGGTATC GGGAAACCGG GTACCGACAA TACCGGCATC 1800
 CTCAATGCTG GCAGCTACAA CACGGGCATC CTCAACGCGG GCGACTTCAA CACGGGCTTC 1860
 TACAACACGG GCACTACAA CACCGGGCGC TTCAACGTCG GTAACACCAA CACCGGCAAC 1920
 TTCAACGTCG GTGACACCAA TACCGGCAGC TATAACCGCG GTGACACCAA CACCGGCTTC 1980
 TTCAATCCCG GCAACGTCAA TACCGGGCGT TCGACACGG GCGACTTCAA CAATGGCTTC 2040
 TTGGTGGCGG GCGATAACCA GGGCCAGATT GCCATCGATC TCTCGGTCAC CACTCCATTC 2100
 ATCCCCATAA ACGAGCAGAT GGTGATTGAC GTACACACCG TAATGACCTT CGGCGGCAAC 2160
 ATGATCAGCG TCACCGAGGC CTCGACGCTT TCGCCGAAA GCTTCTATCT GAGCGGTTTC 2220
 TTCTTCTTGG GCGCGGTCAA TCTCAGCGCA TCCACGCTGA CCGTTCCGAC GATCACCCTC 2280
 ACCATCGGGG GACCGACGGT GACCGTCCC ATCAGCATTG TCGGTGCTCT GGAGAGCCCG 2340
 ACGATTACCT TCCTCAAGAT CGATCGGGCG CCGGGCATCG GAAATTCGAC CACCAACCCC 2400
 TCGTCCGGCT TCTTCAACTC GGGCACCGGT GGCACATCTG GCTTCCAAAA CCGCGGCGGC 2460
 GCGASTTCAG GCGTCTGGAA GATGCTTTTG AGCAGCGCGA TAGGGGAATC GGGTTTCAG 2520
 AACCTCGGCT CGCTSCAGTC AGGCTGGGCG AACCTGGGCA ACTCGGATC GGGCTTTTTT 2580
 AACACAGTA CGGTGAACCT GTGACGCGC GCGAATGCTT CCGGCGTGA CAACATCGGC 2640
 AACACCTCT CCGGCTGCTT CCGGCTGCG AGCGGAGCGA TTTTCAACGC GGGCCTTGGC 2700
 AACCTGGGCG AGTTGAACAT CCGGAGCGCG TCGTCCCGAA TTCGGCACGA GTTAGATACG 2760
 TTTTCAACAA TCATATCGCG GTTTTGGGCG ATGCAATGAG ACGAATCGAA TCGGGGAAGC 2820
 TTAATTAAT AAACCAATC TCGGCTGCTT AT 2880

SEQUENCE INFORMATION FOR SEQ ID NO. 104

- SEQUENCE CHARACTERISTICS
- (A) LENGTH: 948 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS
 - (D) TOPOLOGY: linear

Met Trp Ala Gln Asp Val Ala Ala Met Phe Gly Tyr His Ala Gly Ala
 20 25 30
 Ser Ala Ala Val Ser Ala Leu Thr Pro Phe Gly Gln Ala Leu Pro Thr
 35 40 45
 Val Ala Gly Gly Gly Ala Leu Val Ser Ala Ala Ala Ala Gln Val Thr
 50 55 60
 Thr Arg Val Phe Arg Asn Leu Gly Leu Ala Asn Val Arg Glu Gly Asn
 65 70 75 80
 Val Arg Asn Gly Asn Val Arg Asn Phe Asn Leu Gly Ser Ala Asn Ile
 85 90 95
 Gly Asn Gly Asn Ile Gly Ser Gly Asn Ile Gly Ser Ser Asn Ile Gly
 100 105 110
 Phe Gly Asn Val Gly Pro Gly Leu Thr Ala Ala Leu Asn Asn Ile Gly
 115 120 125
 Phe Gly Asn Thr Gly Ser Asn Asn Ile Gly Phe Gly Asn Thr Gly Ser
 130 135 140
 Asn Asn Ile Gly Phe Gly Asn Thr Gly Asp Gly Asn Arg Gly Ile Gly
 145 150 155 160
 Leu Thr Gly Ser Gly Leu Leu Gly Phe Gly Gly Leu Asn Ser Gly Thr
 165 170 175
 Gly Asn Ile Gly Leu Phe Asn Ser Gly Thr Gly Asn Val Gly Ile Gly
 180 185 190
 Asn Ser Gly Thr Gly Asn Trp Gly Ile Gly Asn Ser Gly Asn Ser Thr
 195 200 205
 Asn Thr Gly Phe Gly Asn Ser Gly Asp Ala Asn Thr Gly Phe Phe Asn
 210 215 220
 Ser Gly Ile Ala Asn Thr Gly Val Gly Asn Ala Gly Asn Tyr Asn Thr
 225 230 235 240
 Gly Ser Tyr Asn Pro Gly Asn Ser Asn Thr Gly Gly Phe Asn Met Gly
 245 250 255
 Gln Tyr Asn Thr Gly Tyr Leu Asn Ser Gly Asn Tyr Asn Thr Gly Leu
 260 265 270
 Ala Asn Ser Gly Asn Val Asn Thr Gly Ala Phe Leu Thr Gly Asn Phe
 275 280 285

305 310 315 320

Asn Ser Gly Ala Gly Ser Ala Ser Gly Phe Leu Asn Ser Gly Ala Asn
325 330 335

Asn Ser Gly Phe Phe Asn Ser Ser Ser Gly Ala Ile Gly Asn Ser Gly
340 345 350

Leu Ala Asn Ala Gly Val Leu Val Ser Gly Val Ile Asn Ser Gly Asn
355 360 365

Thr Val Ser Gly Leu Phe Asn Met Ser Leu Val Ala Ile Thr Thr Pro
370 375 380

Ala Leu Ile Ser Gly Phe Phe Asn Thr Gly Ser Asn Met Ser Gly Phe
385 390 395 400

Phe Gly Gly Pro Pro Val Phe Asn Leu Gly Leu Ala Asn Arg Gly Val
405 410 415

Val Asn Ile Leu Gly Asn Ala Asn Ile Gly Asn Tyr Asn Ile Leu Gly
420 425 430

Ser Gly Asn Val Gly Asp Phe Asn Ile Leu Gly Ser Gly Asn Leu Gly
435 440 445

Ser Gln Asn Ile Leu Gly Ser Gly Asn Val Gly Ser Phe Asn Ile Gly
450 455 460

Ser Gly Asn Ile Gly Val Phe Asn Val Gly Ser Gly Ser Leu Gly Asn
465 470 475 480

Tyr Asn Ile Gly Ser Gly Asn Leu Gly Ile Tyr Asn Ile Gly Phe Gly ..
485 490 495

Asn Val Gly Asp Tyr Asn Val Gly Phe Gly Asn Ala Gly Asp Phe Asn
500 505 510

Ile Gly Phe Ala Asn Thr Gly Asn Asn Asn Ile Gly Phe Ala Asn Thr
515 520 525

Thr Asn Asn Asn Ile Gly Ile Gly Leu Ser Gly Asp Asn Gln Gln Gly
530 535 540

Phe Asn Ile Ala Ser Gly Trp Asn Ser Gly Thr Gly Asn Ser Gly Leu
545 550 555 560

Phe Asn Ser Gly Thr Asn Asn Val Gly Ile Phe Asn Ala Gly Thr Gly
565 570 575

Gly Ile Leu Asn Ala Gly Asp Phe Asn Thr Gly Phe Tyr Asn Thr Gly
 610 615 620
 Ser Tyr Asn Thr Gly Gly Phe Asn Val Gly Asn Thr Asn Thr Gly Asn
 625 630 635 640
 Phe Asn Val Gly Asp Thr Asn Thr Gly Ser Tyr Asn Pro Gly Asp Thr
 645 650 655
 Asn Thr Gly Phe Phe Asn Pro Gly Asn Val Asn Thr Gly Ala Phe Asp
 660 665 670
 Thr Gly Asp Phe Asn Asn Gly Phe Leu Val Ala Gly Asp Asn Gln Gly
 675 680 685
 Gln Ile Ala Ile Asp Leu Ser Val Thr Thr Pro Phe Ile Pro Ile Asn
 690 695 700
 Glu Gln Met Val Ile Asp Val His Asn Val Met Thr Phe Gly Gly Asn
 705 710 715 720
 Met Ile Thr Val Thr Glu Ala Ser Thr Val Phe Pro Gln Thr Phe Tyr
 725 730 735
 Leu Ser Gly Leu Phe Phe Phe Gly Pro Val Asn Leu Ser Ala Ser Thr
 740 745 750
 Leu Thr Val Pro Thr Ile Thr Leu Thr Ile Gly Gly Pro Thr Val Thr
 755 760 765
 Val Pro Ile Ser Ile Val Gly Ala Leu Glu Ser Arg Thr Ile Thr Phe
 770 775 780
 Leu Lys Ile Asp Pro Ala Pro Gly Ile Gly Asn Ser Thr Thr Asn Pro
 785 790 795 800
 Ser Ser Gly Phe Phe Asn Ser Gly Thr Gly Gly Thr Ser Gly Phe Gln
 805 810 815
 Asn Val Gly Gly Gly Ser Ser Gly Val Thr Asn Ser Glu Leu Ser Ser
 820 825 830
 Ala Ile Gly Asn Ser Gly Phe Gln Asn Leu Gly Ser Leu Gln Ser Gly
 835 840 845
 Thr Ala Asn Leu Gly Asn Ser Val Ser Gly Phe Phe Asn Thr Ser Thr
 850 855 860
 Val Asn Leu Ser Thr Pro Ala Asn Val Ser Gly Leu Asn Asn Ile Gly
 865

Ala Gly Leu Ala Asn Leu Gly Gln Leu Asn Ile Gly Ser Ala Ser Cys
 900 905 910

Arg Ile Arg His Glu Leu Asp Thr Val Ser Thr Ile Ile Ser Ala Phe
 915 920 925

Cys Gly Ser Ala Ser Asp Glu Ser Asn Pro Gly Ser Val Ser Glu
 930 935 940

(2) INFORMATION FOR SEQ ID NO:205:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 53 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

XXI. SEQUENCE DESCRIPTION: SEQ ID NO:205:

GGATCCATAT GGGCCATCAT CATCATCATC ACGTGATCGA CATCATCGGG ACC

53

(2) INFORMATION FOR SEQ ID NO:206:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 42 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

XXI. SEQUENCE DESCRIPTION: SEQ ID NO:206:

GCTGAATTCG GGCCTCGGTT GCGCCGCGCCT CATCTTGAAC GA

42

(2) INFORMATION FOR SEQ ID NO:207:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

XXI. SEQUENCE DESCRIPTION: SEQ ID NO:207:

AGATGTTTCA GGTTCGAAAC CAGCGAGCGG T

31

(2) INFORMATION FOR SEQ ID NO:208:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid

CTGTGAATTC AGCGCTGGAA ATCGTCGGCA T

31

(2) INFORMATION FOR SEQ ID NO:209:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:209:

GGATCCAGCG CTGAGATGAA GACCGATGCC GCT

33

(2) INFORMATION FOR SEQ ID NO:210:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:210:

GGATATCTGC AGAATTCAGG TTAAAGGCG ATTTGCCA

38

(2) INFORMATION FOR SEQ ID NO:211:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:211:

GGGATGCGCA GCGACGTGGT CACACGCGCC

30

(2) INFORMATION FOR SEQ ID NO:212:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:212:

CTTATGAAA TTCTAGGCG GTTAAGGTTT TGTGCGG

37

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:213:

TGGCGAATGG GACGCGCCCT GTAGCGGCGC ATTAAGCGCG GCGGGTGTGG TGGTTACGGG	60
CAGCGTGACC GCTACACTTG CCAGCGCCCT AGCGCCCGCT CCTATCGCTT TCTTCCCTTC	120
CTTTCTCGCC ACGTTCGCGG GCTTTCCCGG TCAAGCTCTA AATCGGGGGG TCCCTTTAGG	180
GTTCCGATTT AGTGCTTTAC GGCACCTCGA CCCCCAAAAA CTTGATTAGG GTGATGGTTC	240
ACGTAGTGGG CCATCGCCCT GATAGACGGT TTTTCGCCCT TTGACGTTGG AGTCCACGTT	300
CTTTAATAGT GGACTCTTGT TCCAAACTGG AACAACTCTG AACCTATCTT CGGTCTATTC	360
TTTGGATTTA TAAGGGATTT TGCCGATTTG GGCTATTGGG TTAAAAAATG AGGTGATTTA	420
ACAAAAATTT AACCCGAATT TTAACAAAAT ATTAACGTTT ACAATTTCAG GTGGCACTTT	480
TGGGGGAAAT GTGCGCGGAA CCGCTATTTC TTTATTTTTC TAAATACATT CAAATATGTA	540
TCCGCTCATG AATTAAITCT TAGAAAACT CATCGAGCAT CAAATGAAAC TGCAATTTAT	600
TCATATCAGG ATTATCAATA CCATATTTT GAAAAAGCGG TTTCTGTAAT GAAGGAGAAA	660
ACTCAGCGAG GCAGTTCCAT AGGATGGCAA GATCCTGGTA TCGGTCTGGG ATTCCGACTC	720
GTCCACATC AATACAACCT ATTAATTTCC CCTCTGAAA AATAAGGTTA TCAAGTGAGA	780
AATCAGCATG AGTGACGACT GAATCGGTTG AGAATGGCAA AAGTTTATGC ATTTCTTTCT	840
AJACTTCTTC AACAGGCCAG CCATTACGCT CTCATGAAA ATGCTGCGA TCAACCAAA	900
CGTTATTCAT TCGTGATTGC CCTGAGCGA GACGAAATAC GCGATCGCTG TTAAGGGAC	960
AATTACAAAC AGGAATCGAA TCCACCGGCT TGAGGAACAG TCCAGTGGCA TCAACAATAT	1020
TTTAAATGA ATCAGGATAT TCTTCTAATA TTTGAATG TCTTTTCTG GCGATCGCAG	1080
CTTAAATGA ATCAGGATAT TCTTCTAATA TTTGAATG TCTTTTCTG GCGATCGCAG	1140
TAAATTTCT GAGCCAGTTT AGTCTGACCA TCTTATCTCT AACATGATTC GCAACGCTAC	1200
TTTTTCTAT TTTGAGAAAC AACTCTGGCG CATCGGCTT GCGATAGAAAT GATAGATTC	1260
TGGACCTTAA TTGCCCCGAA TTAATCGGAG GCGATGATA GCGATATAAA TTAGCATGCA	1320
TCTTGAATG GATCGGCTT TTAGGCAAG AGCTTCTTCT TTAGGCAAG	

GATCCTTTT TTCTGCGCGT AATCTGCTGC TTGCAAACAA AAAAACCACC GCTACCAGCG 1560
GTGGTTTGTG TGCCGGATCA AGAGCTACCA ACTCTTTTTC CGAAGGTAAC TGGCTTCAGC 1620
AGAGCGCAGA TACCAAATAC TGTCTTCTA GTGTAGCCGT AGTTAGGCCA CCACTTCAAG 1680
AACTCTGTAG CACCGCCTAC ATACCTCGCT CTGCTAATCC TGTTACCAGT GGCTGCTGCC 1740
AGTGGCGATA AGTCTGTCT TACCGGGTTG GACTCAAGAC GATAGTTACC GGATAAGGCG 1800
CAGCGGTCGG GCTGAACGGG GGGTTCTGTC ACACAGCCCA GCTTGGAGCG AACGACCTAC 1860
ACCGAACTGA GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC CGAAGGGAGA 1920
AAGGCGGACA GGTATCCGGT AAGCGGCAGG GTCCGAACAG GAGAGCGCAC GAGGGAGCTT 1980
CCAGGGGAAA ACCCCTGGTA TCTTTATAGT CCGTCCGGGT TTCCGCACTT CTGACTTGAG 2040
CTCGATTTT TGTGATGCTC CTCAGGGGGG CGGAGCCTAT GGA AAAACGC CAGCAACGCG 2100
GCCTTTTAC GGTTCCTGGC CTTTTGCTG CTTTTCTCTC ACATGTTCTT TCCTGCGTTA 2160
TCCCTGATT CTGTGGATAA CCTATTACC GCCTTTGACT GAGCTGATAC CGCTCGCCGC 2220
AGCCGAACGA CCGAGCGCAG CGAGTCACTG AGCGAGGAAG CGGAAGAGCG CCTGATCGCG 2280
TATTTCTCC TTACCATCT GTCCGCTATT TCACACCGCA TATATGGTGC ACTCTCAGTA 2340
CAATCTGCTC TGATGCGCA TAGTTAAGCC AGTATACCT CCGCTATCGC TACGTGACTG 2400
GCTCTGGCT GCGCGCGCAG ACCCGCAAC ACCCGCTGAC GCGCGCTGAC GGGCTTCTCT 2460
GTCCCGGCA TCCGCTTACA CACAGCTCT CAGCTCTCC CGAGCTGCA TGTGTGAG 2520
TTTTCTCCG TCTATCGCA AACCGCGCAG CAGCTGCGG TAAAGCTCT CAGCTGCTC 2580
CTGAAGCAT TCACAGATCT CTGCTCTTC ATCCGCTCC AGCTGCTTGA GTTCTGCCAG 2640
AACCTTAT GTCTGCTT TCTAAAGCT CCGCTCTTA AGGCGCTTT TTCTCTTT 2700
TCTCTGAT CCTCTCTT AAGGGGATTT TCTCTCTT TCTCTCTT TACCTATGAA 2760
AATGAGAGG ATCTCTGCA TACCGCTTA TCTCTGAA TATCTCTCT TACTGGAACG 2820
TCTGAGCT AAACAAGTCT CCGTATGAT GCGCGCGGAT CAGAGAAAA TCACTCAGGG 2880
TCAATGCGAG CCGTTCTT ATACAGATCT AGCTCTTCA CAGGCTAGCC AGCAGCTCT 2940
TCTGATGAG ATCCGCAAT TCTCTCTT TCTCTCTT TCTCTCTT TCTCTCTT 3000
TCTCTCTT TCTCTCTT TCTCTCTT TCTCTCTT TCTCTCTT TCTCTCTT

CATGCCGGCG ATAATGGCCT GCTTCTCCCG GAAACGTTTG GTGGCGGGAC CAGTGACGAA 3240
GGCTTGAGCG AGGGCGTGCA AGATTCCGAA TACCGCAAGC GACAGGCCGA TCATCGTCGC 3300
GCTCCAGCGA AAGCGGTCCT CGCCGAAAAT GACCCAGAGC GCTGCCGGCA CCGTGCTCTAC 3360
GAGTTGCATG ATAAAGAAGA CAGTCATAAG TGCGGGCAGC ATAGTCATGC CCCGCGCCCA 3420
CCGGAAGGAG CTGACTGGGT TGAAGGCTCT CAAGGGGCATC GGTGAGATC CCGGTGCCTA 3480
ATGAGTGAGC TAACTTACAT TAATTGCGTT GCGCTCACTG CCGGCTTTCC AGTCGGGAAA 3540
CCTGTCTGTC CAGCTGCATT AATGAATCGG CCAACGCGCG GGGAGAGGCG GTTTGCGTAT 3600
TGGGCGCCAG GGTGGTTTTT GTTTTCACCA GTGAGACGGG CAACAGCTCA TTGCCCTTCA 3660
CCGCGTGGCG CTGAGAGACT TCCAGCAAGC GGTCCACGCT GGTTCGCCCC AGCAGGCGAA 3720
AATCCTGTTT GATGGTGGTT AACGGCGGGA TATAACATGA GCTGTCTTCC GTATCGTCCT 3780
ATCCCACTAC CGAGATATCC GCACCAACGC GCACCCCGGA CTGGTAATG GCGCGCATTC 3840
CGCCGAGCGC CATCTGATCG TTGGCAACCA GCATCCGAGT GGGAAACGATG CCGTCATTCA 3900
GCATTTGCAAT GGTTCCTTGA AAACCGGACA TGGCACTCCA GTCCGCTTCC CGTTCCGCTA 3960
TCCGCTGAAT TTGATTGCGA GTGAGATATT TATGCCAGCC AGCCAGACGC AGACGCGCCG 4020
AGACAGAACT TAATGGGCGC GCTAACAGCG CGATTGCTG GTGACCCCAAT GCGACAGAT 4080
GCTCCACGCG CAGTCCGCTA CGCTCTTCTT GGGAGAAAAT AATACTGTTG ATGGGTGTCT 4140
GTTCAGAGAC ATCAGAAAAT AAACCGGAA CATAGTCCA GCGAGCTTCC ACAGCAATGG 4200
CATCCTGGTC ATCCAGCGGA CATTAATGA TCAGCCCACT GACGCGTTGC GCGAGAAGAT 4260
TGTGCACGCG CGCTTTACAG GCTTCGACCG CGCTTCCTTC TACCATCGAC ACCACGCGC 4320
TGTACCGCG TTGATCGCGT CAGATTTAAT TCGCCCGAC AATTTCGCGAC TCGGCTGCA 4380
TGGGCAAT GCAAGTGGCA ATCAGATCA TCAAGCAATG TTTCGCGCG AGTTCTGCTG 4440
CGAGCTGTTT GCGAATGTAA TCAAGTGGC CATCGCGCG TTCCACTTTT TCGCGGTTT 4500
TCCGAGAAAC GTGGCTGGCG TGGTTCGCA TCGCGGAAAC GGTGTGATAA GAGACACCGG 4560
CATACTGTG CAGATCGTAT AACCTTACTG GTTTCAGATC CACCACTCTG AATTGACTCT 4620
CTTCGCGCG GTATCTGCG ATATCGGAA AGCTTTTCA CATTCGATG GTTCGCGCA 4680

CCGGCCACGG GGCCTGCCAC CATACCCACG CCGAAACAAG CGCTCATGAG CCGGAAGTGG 4860
CGAGCCCGAT CTTCGCCATC GGTGATGTGG GCGATATAGG CGCCAGCAAC CGCACCTGTG 4920
GCGCCCGTGA TGCCGGCCAC GATCCGTCCG GCGTAGAGCA TCGAGATCTC GATCCCGCGA 4980
AATTAATACG ACTCACTATA GGGGAATTGT GAGCGGATAA CAATTCGCCCT CTAGAAATAA 5040
TTTTGTTTAA GTTTAAGAAG GAGATATACA TATGGGCCAT CATCATCATC ATCACGTGAT 5100
CGACATCATC GGGACCAGCC CCACATCCTG GGAACAGGCG CCGGCGGAGG CGGTCCAGCG 5160
GGCCCGGGAT AGCGTCGATG ACATCCGCGT CGCTCGGGTC ATTGAGCAGG ACATGGCCGT 5220
GGACAGCCCC GCGAAGATCA CCTACCGCAT CAAGCTCGAA GTGTCTTCA AGATGAGGCC 5280
GCGGCAACCG AGGGGCTGCA AACCACCGAG CGGTTCGCTT GAAACGGGCG CCGGCGCCCG 5340
TACTGTCCCG ACTACCCCGG CGTCGTCCCG GGTGACGTTG GCGGAGACCG GTAGCAGCGT 5400
GCTGTACCCG CTGTTCAACC TGTGGGGTCC GGCCTTTTCA GAGAGGTATC CGAACGTCAC 5460
GATCACCGCT CAGGSCACCG GTTCTGGTGC CGGGATCGCG CAGGCCCGCG CCGGGACCGT 5520
CAACATTGGG GCCTCCGACG CCTATCTGTT GGAAGGTGAT ATGGCCSCCG ACAAGGGGCT 5580
GATGAACATG CGCTAGCCA TCTCGGTCA GAGGTCAAC TACAACCTGC CCGGAGTGAG 5640
CGAGCAGCTC AAGCTGAACG GAAAGTCTT GCGGGCCATG TACCAGGGCA CCATCAAAAC 5700
GTGGGACGAC CCGCAGATCG GTGCGCTCAA GCGCGGCTG AACCTGCGCG GCACCCCGGT 5760
ACTTCGCTG LACCGTCCG ACGGCTCCG TGACGCTTC TTGTTGAGG AGTACCTGTC 5820
CAAGCAAGAT CCGGAGGGCT GGGGCACTG GCGGCTTC GGCACCAAC TCGACTTCCC 5880
GGCGGTGCCG GGTGCGCTGG GTGAGAACCG CAACGGGCGC ATGGTGACCG GTTCCGCGCA 5940
GACACCGGGT TCGTCCGCT ATATCGGCAT GAGCTTCTG TACGAGTCA GTGACCGGG 6000
GTTGGCGAG GTTCACTAG GATATGCTG GAGGATCTG GTTTCGCTG AGCGGAAAG 6060
ATTGAGCGG GCGGCGCTG GTTTCGCTG GAAAGTCTG GCGAAGCAGG GATTTTCGAT 6120
GATGACCGG CCGGCGCGCG ACGGCTACCG GATCATCAAG TACGAGTAC GATGCTCAA 6180
GAGGCGGAA AAGGAGCGCG GCACCGGCA GAGCTTCAG GATTTCTG AGTGGGCGAT 6240
GAGGACCGG AAGGAGCTG GTTTCGCTG GAGGATCTG GTGAGCTG TCGGCGCGG 6300

GGGGACGGCC GCGCAGGCGG CGGTGGTGGG CTTCGAAGAA GCAGCCAAATA AGCAGAAGCA 6540
 GGAACCTGAC GAGATCTCGA CGAATATTCG TCAGGCCCGG GTCCAACTACT CGAGGGGCGA 6600
 CGAGGAGCAG CAGCAGGCGC TGTCCTCGCA AATGGGCTTT GTGCCCACAA CGGCCGCGTC 6660
 GCGCGCGTCG ACCCGTGCAG CGCCACCCGC ACCGGCGACA CTTGTTGCCC CCCCACCACC 6720
 GGCGCGCGCC AACACGCCGA ATGCCAGCC GGGCGATCCC AACGCAGCAC CTCCGCGGGC 6780
 CGACCCGAAC GCACCGCGCG CACCTGTCAT TGCCCCAAAC GCACCCCAAC CTGTCCGGAT 6840
 CGACAACCCG GTTGGAGGAT TCAGCTTCGC GCTGCTGCT GGCTGGGTGG AGTCTGACGC 6900
 CCGCCACTTC GACTACGGTT CAGCACTGCT CAGCAAAACC ACCGGGGACC CGCCATTTCG 6960
 CGGACAGCCG CCGCGCGTGG CCAATGACAC CGGTATCGTG CTCGGCGCGG TAGACCAAAA 7020
 GCTTTACGCC AGCGCCGAAG CCACCGACTC CAAGGCCCGG GCGCGGTTGG GCTCGGACAT 7080
 CGGTGACTTC TATATGCCCT ACCCGGGCAC CCGGATCAAC CAGGAAACCG TCTCGCTTGA 7140
 TGCCAAACGG GTGTCTGGAA GCGCGTCTTA TTACGAAGTC AAGTTCAGCG ATCCGAGTAA 7200
 GCGGAACCGG CAGATCTGGA CCGCGCTAAT CGGCTCGCCC GCGCGGAACG CACCGGACCG 7260
 CCGGCGCGCT CAGCGCTGGT TTGTGGTATG GCTCGGGACC GCGCAACAAC CGGTGGACAA 7320
 GGGCGCGGCC AAGGCGCTGG CGAATCGAT CCGGCTTTG GTGCGCGCGG CGCGGGCGCG 7380
 GGCACCGGCT CCGCAGAGC CGGCTCGCGG GCGCGCGCGG GCGGGGAAG TCGCTCTTAC 7440
 CCGGACGACA CCGACACCGG AGCGGACCTT ACCGCGCTGA GAATTCTGCA CATATCCATC 7500
 ACACTGGCGG CCGCTCGAGC ACCACACCA CGACCACTGA GATCGCGCTG CTACAAAGC 7560
 CCGAAAGGAA GTCAGTTGG CTCTTCGCA CTCTTACCA TAACTAGCAT AACCGCTTGG 7620
 GCTCTTAAA GCTCTTGA GCGCTTTTCT GCTGAAGGA TAACTATAT CGGAT 7680

INFORMATION FOR SEQ ID NO:114.

SEQUENCE CHARACTERISTICS.

- A LENGTH: 802 amino acids
- P TYPE: amino acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

11 SEQUENCE DESCRIPTION: (b) (4).

206

20 25 30
 Asp Ser Val Asp Asp Ile Arg Val Ala Arg Val Ile Glu Gln Asp Met
 35 40 45
 Ala Val Asp Ser Ala Gly Lys Ile Thr Tyr Arg Ile Lys Leu Glu Val
 50 55 60
 Ser Phe Lys Met Arg Pro Ala Gln Pro Arg Gly Ser Lys Pro Pro Ser
 65 70 75 80
 Gly Ser Pro Glu Thr Gly Ala Gly Ala Gly Thr Val Ala Thr Thr Pro
 85 90 95
 Ala Ser Ser Pro Val Thr Leu Ala Glu Thr Gly Ser Thr Leu Leu Tyr
 100 105 110
 Pro Leu Phe Asn Leu Trp Gly Pro Ala Phe His Glu Arg Tyr Pro Asn
 115 120 125
 Val Thr Ile Thr Ala Gln Gly Thr Gly Ser Gly Ala Gly Ile Ala Gln
 130 135 140
 Ala Ala Ala Gly Thr Val Asn Ile Gly Ala Ser Asp Ala Tyr Leu Ser
 145 150 155 160
 Glu Gly Asp Met Ala Ala His Lys Gly Leu Met Asn Ile Ala Leu Ala
 165 170 175
 Ile Ser Ala Gln Gln Val Asn Tyr Asn Leu Pro Gly Val Ser Glu His
 180 185 190
 Leu Lys Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile
 195 200 205
 Lys Thr Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn
 210 215 220
 Leu Pro Gly Thr Ala Val Val Pro Leu His Arg Ser Asp Gly Ser Gly
 225 230 235 240
 Asp Thr Phe Leu Thr Thr Gln Pro Leu Ser Lys Gln Asp Pro Glu Gly
 245 250 255
 Trp Gly Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val
 260 265 270
 Pro Gly Ala Leu Gly His Asn Gly Asn Gly Gly Met Val Thr Gly Cys
 275 280 285 290 295
 Ala Gly Thr ...

Gly Asn Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala
325 330 335

Gly Phe Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp
340 345 350

Gly Pro Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile
355 360 365

Val Asn Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala
370 375 380

Phe Leu His Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp
385 390 395 400

Gln Val His Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp
405 410 415

Ala Leu Ile Ala Thr Ile Ser Ser Ala Glu Met Lys Thr Asp Ala Ala
420 425 430

Thr Leu Ala Gln Glu Ala Gly Asn Phe Glu Arg Ile Ser Gly Asp Leu
435 440 445

Lys Thr Gln Ile Asp Gln Val Glu Ser Thr Ala Gly Ser Leu Gln Gly
450 455 460

Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln Ala Ala Val Val Arg
465 470 475 480

Phe Gln Glu Ala Ala Asn Lys Gln Lys Gln Glu Leu Asp Glu Ile Ser
485 490 495

Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg Ala Asp Glu Glu
500 505 510

Gln Gln Gln Ala Leu Ser Ser Gln Met Gly Phe Val Pro Thr Thr Ala
515 520 525

Ala Ser Pro Pro Ser Thr Ala Ala Ala Pro Ala Pro Ala Thr Pro
530 535

Val Ala Pro Pro Pro Pro Ala Ala Ala Asn Thr Pro Asn Ala Gln Pro
545 550 555 560

Gly Asp Pro Asn Ala Ala Pro Pro Pro Ala Asp Pro Asn Ala Pro Pro
565 570 575

Pro Pro Val Ile Ala Pro Asn Ala Pro Gln Pro Val Arg Thr Thr
580 585

Asp Ala Ala His Phe Asp Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr
 610 615 620
 Gly Asp Pro Pro Phe Pro Gly Gln Pro Pro Pro Val Ala Asn Asp Thr
 625 630 635 640
 Arg Ile Val Leu Gly Arg Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu
 645 650 655
 Ala Thr Asp Ser Lys Ala Ala Ala Arg Leu Gly Ser Asp Met Gly Glu
 660 665 670
 Phe Tyr Met Pro Tyr Pro Gly Thr Arg Ile Asn Gln Glu Thr Val Ser
 675 680 685
 Leu Asp Ala Asn Gly Val Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys
 690 695 700
 Phe Ser Asp Pro Ser Lys Pro Asn Gly Gln Ile Trp Thr Gly Val Ile
 705 710 715 720
 Gly Ser Pro Ala Ala Asn Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp
 725 730 735
 Phe Val Val Trp Leu Gly Thr Ala Asn Asn Pro Val Asp Lys Gly Ala
 740 745 750
 Ala Lys Ala Leu Ala Glu Ser Ile Arg Pro Leu Val Ala Pro Pro Pro
 755 760 765
 Ala Pro Ala Pro Ala Pro Ala Glu Pro Ala Pro Ala Pro Ala
 770 775 780
 Gly Glu Val Ala Pro Thr Pro Thr Thr Pro Thr Pro Gln Arg Thr Leu
 785 790 795 300
 Pro Ala

INFORMATION FOR SEC ID NO. 115.

SEQUENCE CHARACTERISTICS

- LENGTH: 484 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

11 MOLECULE TYPE: Genomic DNA

11 SEQUENTIAL DESCRIPTION: SEC. 10. 11. 12. 13.

270 280 290 300 310 320 330 340 350 360 370 380 390 400 410 420 430 440 450 460 470 480 490 500 510 520 530 540 550 560 570 580 590 600 610 620 630 640 650 660 670 680 690 700 710 720 730 740 750 760 770 780 790 800 810 820 830 840 850 860 870 880 890 900 910 920 930 940 950 960 970 980 990

CCGATCGCGA TCAGCTGCTT ACCGACCGGC GGGTGAACCA CCAGGCCGTA CCCGGGGTTG 360
 TCTTCCACCC CATGGTTGTT CAGCACCTGC CAGGCCTGGC GGTGCGTAAT GCTTCTCGTC 420
 GAAGATGGGG GTGCCCGCAT CCGTCACCGA GCCC 454

(2) INFORMATION FOR SEQ ID NO:216:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 470 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:216:

TCCAGAAGTA CGGCGGATCC TCGGTGGCCG AC3CCGAACG GATTCCGCCG GTC3CCGAAC 50
 GCATCGTCCG CACCAAGAAG CAAGGCAATG ACCTCGTCTG CTTCTCTCTT GCCATGGGGG 100
 ATACCACCGA CGACCTGCTG GATCTGGCTG AGCAGGTGTG CCGCGCCCGG CCGCCTCGGG 150
 AGCTGGACAT GCTGCTTACC GCCGGTGAAC GCATCTCGAA TCGGTTGGTG GCCATGGCCA 200
 TCGAGTCGCT CGGCGCGCAT GCCCGGTCTG TCACCGGTTT GCAGGCCGGG GTGATCACCA 240
 CCGGCACCCA CGGCAACGCC AAGATCATCG ACGTCACGCC GGGGCGGGTG CAAACCGCCC 300
 TTGAGGAAGG GCGGGTCGTC TTGGTGGCCG GATTCCAAGG GGTGAGCCAG GACACCAAGG 360
 ATGTCACGAC GTTGGGCCCG GCGGCTCCG ACACCCACCG CGTCGCCATG 420
 470

(2) INFORMATION FOR SEQ ID NO:217:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 279 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:217:

TCCCGGCGTA CCGGCGCGCG ACAAACAACG ATCGATTGAT ATCGATGAGA GACCGAGGAA 50
 TCTTGGGCTT TCCCGAGTTC ACCGACGAGT AGCGCGCGCG CGCCTTGGAG AAGGCTGCTG 100
 TTTACGCTCG AGCGCGAGCA GAGCTCAAGG ATCGGCTGAA TCGTGGCGGC ACCAAGCTCA 150
 TTTAGTCTCT CAAGGACCGG GAGAGCGATG AGTCTTTGGG TAAATGTAAG TTGTCTGCGT 200
 TTTCTTAGGG GTTGGCAAGG TTGGCAAGG TCCAGGCGC 250

(2) INFORMATION FOR SEQ ID NO:218:

(i) SEQUENCE CHARACTERISTICS:

- A. LENGTH: 219 base pairs
 B. TYPE: nucleic acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

ACACGGTCCA	ACTCGACGAG	CCCCCTGCTGG	AGGTGTGCGAC	CGACAAGGTC	GACACCGAAA	60
TCCCTCGCCG	GCCGCGGGTG	TGCTGACCAA	GATCATCGCC	CAAGAAGATG	ACACGGTCCA	120
GGTCGGCGGC	GAGCTCTCTG	TCATTGGCGA	CGCCCATGAT	GCCGGCGAGG	CCGCGGTCCC	180
GGCACCCGAG	AAAGTCTCTG	CCGGCCCCAAC	CCGAATCCA			219

(2) INFORMATION FOR SEQ ID NO:219:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 342 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:219:

TGCTGCGCGA	GATCGGCGCC	GGCGGCGCGC	CCAAGCCCGC	ACCCAAGCCC	GTCCCCGAGC	60
CAGCGCCGAC	GCCGAAGGCC	GAACGCGCAC	CATCGCCGCG	GGCGGCGCGC	CCAGCCCGGTG	120
CGGCGGAGGG	CGCACCGTAC	GTGAAGCCGC	TGGTGCGAAA	GCTGGCGTCG	GAAAACAACA	180
TGACCTCGC	CGGGGTGACC	GGCACCGGAG	TGGGTGGTCC	CATCCGCAAA	CAGGATGTGC	240
TGGCGCGCGC	TGAACAAAAG	AAGCGGCGCA	AAGCACCGGC	GCCGGCGCGC	CAGGCGCGCC	300
CCGCGCGCGC	CCCGAAAGCG	CCGCTGAAG	ATCCGATGCC	GC		342

(2) INFORMATION FOR SEQ ID NO:220:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 515 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:220:

GGTCTTGGT	CAGTATCAGC	GGGACGAGG	AGGCGACGGT	GGCGTCCGC	GGCGAGTTGG	60
TTGGGATCGG	TGTGGCTGCC	GACATCGGTG	GGCGGCGCGC	CCCAAGCCC	GCACCCAAGC	120
TTGTCTCTGA	GCGAGCGCCG	ACCGGGAAGG	GGAAACGCGC	AGGATCGCGC	CGGGGGGCGC	180
ATGATCTGG	TGGGGCGGAG	GGCGACGCT	AGGTGACGCG	GCTGGTGGCA	AAGGTGGCGT	240
AGAAAACAA	ATGAGACCTC	GGGGGCTGA	CGGCGACGCG	AGTGGGTGGT	GGCATCGGCA	300
AATAGGATGT	GCTGGCGCGC	GCTGAACAAA	AGAAGCGCGC	GAAAGCACCG	GGGGGCTGAG	360
AGGTCATCA	CGGGTTAAG	CAGTTTCCCG	GAGAAGCGCG	GTTGACCTC	TTGGGGGCTC	420
TTGGTCTGGT	CGAGGCGGTC	GGGAGCGGAG	TTGAGGTTAG	GGGGGCGAAA	TCTTCAAGTT	480
GGGAGGGAAG	GGCACTCGGA	AGAGGCTCGC	CACCC			515

(2) INFORMATION FOR SEQ ID NO:221:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 515 base pairs
- (B) TYPE: nucleic acid

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:221:

CCGACCCCAA	GGTGCAGATT	CAACAGGCCA	TTGAGGAAGC	ACAGCGCACC	CACCAAGCGC	60
TGACTCAACA	GGCGGCGCAA	GTGATCGGTA	ACCAGCGTCA	ATTGGAGATG	CGACTCAACC	120
GACAGCTGGC	GGACATCGAA	AAGCTTCAGG	TCAATGTGCG	CCAAGCCCTG	ACGCTGGCCG	180
ACCAGGCCAC	CGCCGCCGGA	GACGCTGCCA	AGGCCACCGA	ATACAACAAC	GCCGCCGAGG	240
CGTTCGCAGC	CCAGCTGGTG	ACCGCCGAGC	AGAGCGTCGA	AGACCTCAAG	ACGCTGCATG	300
ACCAGGCGCT	TAGCGCCGCA	GCTCAGGCCA	AGAAGGCCGT	CGAACGAAAT	GCGATGGTGC	360
TGCAGCAGAA	GATCGCCGAG	CGAACCAAGC	TGCTCAGCCA	GCTCGAGCAG	GCGAAGATGC	420
AGGAGCAGGT	CAGCGCATCG	TTGCGGTCGA	TGAGTGAGCT	CGCCGCGCCA	GGCAACACGC	480
CGAGCCTCGA	CGAGGTGCGC	GACAAGATCG	AGCGTCGCTA	CGCCAACGCG	ATCGGTTCCG	540
CTGAACCTTG	CGAGAGT					557

(2) INFORMATION FOR SEQ ID NO:222:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 223 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:222:

CAGGATAGGT	TTCCACATCC	ACCTGGGTTG	CGCAGCCGGT	GCSCGACCGT	GTGATAGGCC	60
AGAGGTGGAC	CTGCCCCGAC	CGACGATCGA	TCCAGGAGTC	AACAGAAATG	GCCTTCTCCG	120
TCCAGATGCC	GGCACTCGGT	GAGAGCGTCA	CGAGGGGAGC	GGTACCCGCG	TGGCTCAAAAC	180
AGGAAGGCCA	CACGGTCCAA	CTCGACGAGT	CCCTCTGGGA	GGT		223

(2) INFORMATION FOR SEQ ID NO:223:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 579 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:223:

AAGAAGTACA	TCTGCGGTTT	GAGGTGCGCG	AAGCAAGGGA	CTCAAGCGGC	GCAGTAGCCG	60
ACCAGGAGCA	CCGCATAAGG	CGAGTCCCGG	CGCAGAAAGA	TAGGCGAGCC	CGGCTATGCC	120
AGGACTGGCA	CGCCGAGGTA	CGACATGCGG	CGCGTCCGTA	CGAGCATGTC	GGCCTTGACG	180
CAGCACTGTG	CGCCGCGAGT	TGCACCTGTT	TGCTGGTGGT	TGCGGTAGAG	CACCGGCCGC	240
AACGACATGG	GCCAGGTGCA	CGCTTTGGAT	TGCAAGGGGT	GGTAGTTGCG	TGCGGAATTC	300
GTGAGGCGCG	CGTGAAGATG	GAACGCTTTG	GGGTGTATTT	CGTAGAGGGA	GCGCACGCGG	360
TCCGCGAGCG	GAACCAAGTA	CTTGGAATCG	ACCGCTTGA	CGAAGGATG	CGGATCGATG	420
CGGCTCTCGG	ACCGCAAGCG	CGAGGATGCG	CGGCTCTCGG	CGGCTCTCGG	CGGCTCTCGG	480

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 484 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:224:

ACAACGATCG ATTGATATCG ATGAGAGACG GAGGAATCGT GGCCCTTCCC CAGTTGACCG	60
ACGAGCAGCG CCGGGCCCGG TTGGAGAAGG CTGCTGCCGC ACCTCGAGCG CGAGCAGAGC	120
TCAAGGATCG GCTCAAGCGT GGCGGCACCA ACCTCACCCA GGTCTCAAG GACGCGGAGA	180
GCGATGAAGT CTTGGGCAAA ATGAAGGTGT CTGCGCTGCT TGAGGCCTTG CCAAAGGTGG	240
GCAAGGTCAA GGCGCAGGAG ATCATGACCG AGCTGGAAAT TGCGGCCAC CCGCGCGCT	300
TCGTGGCCTC GGTCAACGTC AGGCAAGGC CCTGCTGGAA AAGTTCGGT CCGCCTAACC	360
CCCGCGGCGG ACGATCGCGG CCGGAAGGCC TGTGGTGGG GTACCCCGCG ATACGGGGGA	420
GAAGCGGCCT GACAGGGCCA GGTCACAATT CAGGCCGAAC GCGCCGCTGG GGGGGAACCC	480
GCCC	484

(i) INFORMATION FOR SEQ ID NO:225:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 537 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:225:

AGGACTGGCA CCGCCAGCCA CCACATCCCG GCGGTGCCGA CCAGCATCTC GGCTTGACG	60
CACGACTGTC CCGCGCAGGC TGCAACGTCT TGCTGGTCCA TGGCGTACAG CACCGGCCGC	120
AACGACATGG CCGAGGTCCA CGGTTTGGAT TCCCAAGGGT GGTAGTTGCC TCGGGAATTC	180
GTGAGGCCCG CCGTGAAGTG GAACGCTTTG GCGGTGTAGT CCGAGAGCGA GCGCACGGCG	240
TGGGGCAGCG GAACAACCGA GTTGGGACCG ACCGCTTGAC CGACCGCATG CCGATCGATC	300
CCGCTCTCGG ACGCGAACCA CGGAGCGTAG GTGGCCAGAT AGACCGCGAA TGGGATCAAC	360
CCGAGCGCAT ACCCGCTGGG AAGGCGCTCA CGCGGCACTG TCCCGAGCCA CGGTCTTTGC	420
ACTTGGTACT GAGTCCGCG CCGACGCTCG AAGGCGAGCG CCATCCGCGG GAAGAACAGC	480
ACGAATACA CCGCGGACCA TTGCTGCGG CAAGCGAATC CCAAGCGCA CCGCGG	537

(i) INFORMATION FOR SEQ ID NO:226:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 135 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

213

Gly Gly Ala Ala Ala Gly Gln Gln Ser Asp Val His Pro Phe Ala Asn
 1 5 10 15
 Leu Ile Ala Val Asp Asp Glu Arg Ala Glu Arg Arg Asp Asp Glu Glu
 20 25 30
 Arg Gln Glu Ala Val Gln Gln Arg Gly Pro Arg Gly Asp Glu Ala Asp
 35 40 45
 Pro Val Ala Asp Gln Gln His Pro Gly Asp Gly Ala Asp Gln Cys Arg
 50 55 60
 Pro Ala Asp Pro Pro His Asp Pro His His Gln Arg His Gln Asp His
 65 70 75 80
 Thr Gln Gln Gly Ala Gly Glu Pro Pro Ala Glu Ser Val Val Thr Glu
 85 90 95
 Asp Gly Leu Pro Asp Arg Asp Gln Leu Leu Thr Asp Arg Arg Val Asn
 100 105 110
 His Gln Ala Val Pro Gly Val Val Phe His Pro Met Val Val Gln His
 115 120 125
 Leu Pro Gly Leu Ala Val Arg
 130 135

(2) INFORMATION FOR SEQ ID NO:227:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 156 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:227:

Gln Lys Tyr Gly Gly Ser Ser Val Ala Asp Ala Glu Arg Ile Arg Arg
 1 5 10 15
 Val Ala Glu Arg Ile Val Ala Thr Lys Lys Gln Gly Asn Asp Val Val
 20 25 30
 Val Val Val Ser Ala Met Gly Asp Thr Thr Asp Asp Leu Leu Asp Leu
 35 40 45
 Ala Gln Gln Val Cys Pro Ala Pro Pro Pro Arg Glu Leu Asp Met Leu
 50 55 60
 Leu Thr Ala Gly Gln Arg Ile Ser Asn Ala Leu Val Ala Met Ala Ile
 65 70 75 80
 His Ser Leu Gly Ala His Ala Arg Ser Phe Thr Gly Ser Gln Ala Gly
 85 90 95
 Val Ile Thr Thr Gly Thr His Gly Asn Ala Lys Ile Ile Asp Val Thr
 100 105 110
 Pro Gly Arg Leu Gln Thr Ala Leu Glu Gln Gly Arg Val Val Leu Val
 115 120 125
 Ala Gly Phe Gln Gly Val Ser Gln Asp Thr Lys Asp Val Thr Thr Leu
 130 135 140
 Gly Arg Gly Gly Ser Asp Thr Thr Ala Val Ala Met
 145

214

- (A) LENGTH: 92 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:228:

```

Pro Ala Tyr Pro Ala Gly Thr Asn Asn Asp Arg Leu Ile Ser Met Arg
 1           5           10           15
Asp Gly Gly Ile Val Ala Leu Pro Gln Leu Thr Asp Glu Gln Arg Ala
      20           25           30
Ala Ala Leu Glu Lys Ala Ala Ala Arg Arg Ala Arg Ala Glu Leu
      35           40           45
Lys Asp Arg Leu Lys Arg Gly Gly Thr Asn Leu Thr Gln Val Leu Lys
      50           55           60
Asp Ala Glu Ser Asp Glu Val Leu Gly Lys Met Lys Val Ser Ala Leu
      65           70           75           80
Leu Glu Ala Leu Pro Lys Val Gly Lys Val Gln Ala
      85           90

```

(2) INFORMATION FOR SEQ ID NO:229:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 72 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:229:

```

Thr Val Glu Leu Asp His Pro Leu Val Glu Val Ser Thr Asp Lys Val
 1           5           10           15
Asp Thr Glu Ile Pro Ser Pro Ala Ala Gly Val Leu Thr Lys Ile Ile
      20           25           30
Ala Gln Glu Asp Asp Thr Val Glu Val Gly Gly Glu Leu Ser Val Ile
      35           40           45
His Asp Ala His Asp Ala Gly Glu Ala Ala Val Pro Ala Pro Gln Lys
      50           55           60
Val Ser Ala Gly Pro Thr Arg Ile
      65           70

```

(2) INFORMATION FOR SEQ ID NO:230:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 113 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:230:

Ala Ala Asp Ile Gly Ala Ala Pro Ala Pro Lys Pro Ala Pro Lys Pro
 1 5 10 15
 Val Pro Glu Pro Ala Pro Thr Pro Lys Ala Glu Pro Ala Pro Ser Pro
 20 25 30
 Pro Ala Ala Gln Pro Ala Gly Ala Ala Glu Gly Ala Pro Tyr Val Thr
 35 40 45
 Pro Leu Val Arg Lys Leu Ala Ser Glu Asn Asn Ile Asp Leu Ala Gly
 50 55 60
 Val Thr Gly Thr Gly Val Gly Gly Arg Ile Arg Lys Gln Asp Val Leu
 65 70 75 80
 Ala Ala Ala Glu Gln Lys Lys Arg Ala Lys Ala Pro Ala Pro Ala Ala
 85 90 95
 Gln Ala Ala Ala Pro Ala Pro Lys Ala Pro Pro Glu Asp Pro Met
 100 105 110
 Pro

(2) INFORMATION FOR SEQ ID NO:231:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 118 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:231:

Val Leu Val Ser Ile Ser Ala Asp Glu Asp Ala Thr Val Pro Val Gly
 1 5 10 15
 Gly Glu Leu Ala Arg Ile Gly Val Ala Ala Asp Ile Gly Ala Ala Pro
 20 25 30
 Ala Pro Lys Pro Ala Pro Lys Pro Val Pro Glu Pro Ala Pro Thr Pro
 35 40 45
 Lys Ala Glu Pro Ala Pro Ser Pro Pro Ala Ala Gln Pro Ala Gly Ala
 50 55 60
 Ala Glu Gly Ala Pro Tyr Val Thr Pro Leu Val Arg Lys Leu Ala Ser
 65 70 75 80
 Ile Asn Asn Ile Asp Leu Ala Gly Val Thr Gly Thr Gly Val Gly Gly
 85 90 95
 Arg Ile Arg Lys Gln Asp Val Leu Ala Ala Ala Glu Gln Lys Lys Arg
 100 105 110
 Ala Lys Ala Pro Ala Pro
 115

(2) INFORMATION FOR SEQ ID NO:231:

SEQUENCE

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:232:

```

Asp Pro Lys Val Gln Ile Gln Gln Ala Ile Glu Glu Ala Gln Arg Thr
 1           5           10           15
His Gln Ala Leu Thr Gln Gln Ala Ala Gln Val Ile Gly Asn Gln Arg
 20           25           30
Gln Leu Glu Met Arg Leu Asn Arg Gln Leu Ala Asp Ile Glu Lys Leu
 35           40           45
Gln Val Asn Val Arg Gln Ala Leu Thr Leu Ala Asp Gln Ala Thr Ala
 50           55           60
Ala Gly Asp Ala Ala Lys Ala Thr Glu Tyr Asn Asn Ala Ala Glu Ala
 65           70           75           80
Phe Ala Ala Gln Leu Val Thr Ala Glu Gln Ser Val Glu Asp Leu Lys
 85           90           95
Thr Leu His Asp Gln Ala Leu Ser Ala Ala Gln Ala Lys Lys Ala
100           105           110
Val Glu Arg Asn Ala Met Val Leu Gln Gln Lys Ile Ala Glu Arg Thr
115           120           125
Lys Leu Leu Ser Gln Leu Glu Gln Ala Lys Met Gln Glu Gln Val Ser
130           135           140
Ala Ser Leu Arg Ser Met Ser Glu Leu Ala Ala Pro Gly Asn Thr Pro
145           150           155           160
Ser Leu Asp Glu Val Arg Asp Lys Ile Glu Arg Arg Tyr Ala Asn Ala
165           170           175
Ile Gly Ser Ala Glu Leu Ala Glu Ser
180           185

```

(2) INFORMATION FOR SEQ ID NO:233:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 71 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(1) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:233:

```

Val Ser Thr Ser Thr Trp Val Pro His Pro Val Arg Asp Arg Val Ile
 1           5           10           15
Gly Gln Arg Trp Thr Cys Ala Asp Arg Arg Ser Ile Glu Glu Ser Thr
 20           25           30
Gln Met Ala Phe Ser Val Gln Met Pro Ala Leu Gly Glu Ser Val Thr
 35           40           45
Gln Gly Thr Val Thr Arg Trp Leu Asp Gln Gln Glu Asp Thr Val Gln
 50           55           60
Leu Asp Gln Ser Val Val Val Val Val Val Val Val Val Val Val Val
 65           70           75           80

```

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 182 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:234:

Glu Val His Leu Pro Val Asp Val Gly Glu Pro Arg Gln Pro Thr Gly
 1 5 10 15
 Ala Val Ala Asp Gln Asp His Arg Ile Thr Pro Val Pro Ala His Lys
 20 25 30
 His Thr Pro Pro Arg Val Cys Gln Asp Trp His Arg Gln Pro Pro His
 35 40 45
 Arg Gly Arg Ala Asp Gln His Leu Gly Leu Asp Ala Arg Leu Cys Ala
 50 55 60
 Ala Ala Cys Asn Val Leu Leu Val Asp Gly Val Gln His Arg Pro Gln
 65 70 75 80
 Arg His Gly Pro Gly Pro Arg Phe Gly Phe Pro Arg Val Val Val Ala
 85 90 95
 Cys Gly Ile Arg Gln Ala Arg Val Gln Val Glu Arg Phe Gly Gly Val
 100 105 110
 Leu Pro Glu Arg Ala His Gly Val Gly Gln Arg Asn Asn Arg Val Ala
 115 120 125
 Thr Asp Arg Leu Thr Asp Arg Met Pro Ile Asp Arg Gly Leu Gly Arg
 130 135 140
 Glu Pro Arg Ser Val Gly Gly Gln Ile Asp Arg Glu Arg Asp Gln Pro
 145 150 155 160
 Gln Arg Ile Pro Ala Gly Cys His Val Thr Pro His Cys Ser Gln Pro
 165 170 175
 Arg Ser Leu His Leu Val
 180

2. INFORMATION FOR SEQ ID NO:235:

(i) SEQUENCE CHARACTERISTICS:

- A. LENGTH: 160 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:235:

Asn Asp Arg Leu Ile Ser Met Arg Asp Gly Gln Ile Val Ala Leu Pro
 1 5 10 15
 Gln Leu Thr Asp Gln Gln Arg Ala Ala Ala Leu Ile Asp Ala Ala Ala
 20 25 30 35 40

Gly Lys Met Lys Val Ser Ala Leu Leu Glu Ala Leu Pro Lys Val Gly
 65 70 75 80
 Lys Val Lys Ala Gln Glu Ile Met Thr Glu Leu Glu Ile Ala Pro His
 85 90 95
 Pro Ala Ala Phe Val Ala Ser Val Thr Val Ser Ala Arg Pro Cys Trp
 100 105 110
 Lys Ser Ser Ala Pro Pro Asn Pro Ala Gly Arg Arg Cys Gly Pro Glu
 115 120 125
 Gly Leu Trp Trp Ala Tyr Pro Arg Ile Arg Gly Arg Ser Gly Leu Thr
 130 135 140
 Gly Pro Ala His Asn Ser Gly Arg Thr Pro Arg Trp Gly Gly Thr Arg
 145 150 155 160

(2) INFORMATION FOR SEQ ID NO:236:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 178 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:236:

Asp Trp His Arg Gln Pro Pro His Arg Gly Arg Ala Asp Gln His Leu
 1 5 10 15
 Gly Leu Asp Ala Arg Leu Cys Ala Ala Cys Asn Val Leu Leu Val
 20 25 30
 Asp Gly Val Gln His Arg Pro Gln Arg His Gly Pro Gly Pro Arg Phe
 35 40 45
 Gly Phe Pro Arg Val Val Val Ala Cys Gly Ile Arg Gln Ala Arg Val
 50 55 60
 Glu Val Glu Arg Phe Gly Gly Val Val Pro Glu Arg Ala His Gly Val
 65 70 75 80
 Gly Gln Arg Asn Asn Arg Val Ala Thr Asp Arg Leu Thr Asp Arg Met
 85 90 95
 Pro Ile Asp Arg Gly Leu Gly Arg Glu Pro Arg Ser Val Gly Gly Gln
 100 105 110
 Ile Asp Arg Glu Arg Asp Gln Pro Gln Arg Ile Phe Ala Gly Lys His
 115 120 125
 Val Thr Pro His Cys Pro Gln Pro Arg Ser Leu His Leu Val Leu Thr
 130 135 140
 Ser Arg Arg His Val Glu Arg Gln Arg His Arg Ala Glu Glu Gln His
 145 150 155 160
 Glu Val His Ala Gly Pro Leu Gly Gly Ala Ser Gln Ser Gln Ala Ala
 165 170 175
 Pro Arg

- (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:237:

```

ATGCCAAGCC GGTGCTGATG CCCGAGCTCG GCGAATCGGT GACCGAGGGG ACCGTCATTC      60
GTGGGCTGAA GAAGATCGGG GATTCGGTTC AGGTTGACGA GCCACTCGTG GAGGTGTCCA      120
CCGACAAGGT GGACACCGAG ATCCCGTCCC CCGTGGCTGG GGTCTTGGTC AGTATCAGCG      180
CCGACGAGGA CGCCACGGTG CCCGTCGGCG GCGAGTTGGC CCGGATCGGT GTCGCTGCCG      240
AGATCGGCGC CGCGCCCGCC CCCAAGCCCC C
                                                                    271
  
```

(2) INFORMATION FOR SEQ ID NO:238:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 89 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:238:

```

Ala Lys Pro Val Leu Met Pro Glu Leu Gly Glu Ser Val Thr Glu Gly
 1           5           10           15
Thr Val Ile Arg Trp Leu Lys Lys Ile Gly Asp Ser Val Gln Val Asp
      20           25           30
Glu Pro Leu Val Glu Val Ser Thr Asp Lys Val Asp Thr Glu Ile Pro
      35           40           45
Ser Pro Val Ala Gly Val Leu Val Ser Ile Ser Ala Asp Glu Asp Ala
      50           55           60
Thr Val Pro Val Gly Gly Glu Leu Ala Arg Ile Gly Val Ala Ala Glu
      65           70           75           80
Ile Gly Ala Ala Pro Ala Pro Lys Pro
      85
  
```

(2) INFORMATION FOR SEQ ID NO:239:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 101 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:239:

TAGTAGCGG ATGCGCGGA ...

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 339 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:240:

ATGAAGTTGA AGTTTGCTCG CCTGAGTACT GCGATACTGG GTTGTGCAGC GCGCGTTGTG 60
 TTTCCTGCCT CGGTTGCCAG CGCAGATCCA CCTGACCCGC ATCAGCCGGA CATGACGAAA 120
 GGCTATTGCC CGGGTGGCCG ATGGGGTTTT GCGCACTTGG CCGTGTGCGA CGGCGAGAAG 180
 TACCCCGACG SCTCGTTTTG GCACCACTGG ATGCAAACGT GGTTTACCGG CCCACAGTTT 240
 TACTTCGATT GTGTCAGCGG CCGTGAGCCC CTCGCCGCGC CGCCGCCACC GGGTGGTTGC 300
 GGTGGGGCAA TTCCGTCCGA GCAGCCCAAC GTCCTCTGA 339

(i) INFORMATION FOR SEQ ID NO:241:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 112 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:

Met Lys Leu Lys Phe Ala Arg Leu Ser Thr Ala Ile Leu Gly Cys Ala
 1 5 10 15
 Ala Ala Leu Val Phe Pro Ala Ser Val Ala Ser Ala Asp Pro Pro Asp
 20 25 30
 Pro His Gln Pro Asp Met Thr Lys Gly Tyr Cys Pro Gly Gly Arg Trp
 35 40 45
 Gly Phe Gly Asp Leu Ala Val Tyr Asp Gly Glu Lys Tyr Pro Asp Gly
 50 55 60
 Ser Phe Trp His Gln Trp Met Gln Thr Trp Phe Thr Gly Pro Gln Phe
 65 70 75 80
 Tyr Phe Asp Tyr Val Ser His Gly Gln Pro Leu Pro Gly Pro Pro
 85 90 95
 Pro Gly Gly Tyr Gly Gly Ala Ile Pro Ser Gln Gln Pro Asp Ala Pro
 100 105 110

(i) INFORMATION FOR SEQ ID NO:242:

(1) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 371 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

```

GTGACCACGG TGGGCGTGGC ACCAACCCGG GCAGCGGCAG CCGCGGCGGG GCCGCGGGCT 60
CCGGCGGCCAA CGGTGGCGGC GGGGGTAACG CCACCGGCTC AGGCGGCAAG GGCGGCGCG 120
GTGGCAATGG CGGTGATGGG AGCTTCGGCG CTACCAAGCG CCGCGCCTCC ATCGGGGTCA 180
CGGGCGCCCC CGGCGGCAAC GCGGCAAGG GCGGCGCCCG TGGCAGCAAC CCGAACGGCT 240
CAGGTGGCGA CGGCGGCAAA GCGGCAACG GCGGTGCCCG CCGCAACGGG GGCTCGATCG 300
GCGCCAACAG CGGCATCGTC GCGGTTCCG GTGGGCGCGG TGGCGCTGGC GGCGCCGCG 360
GAAACGGCAG C

```

(2) INFORMATION FOR SEQ ID NO:243:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 424 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:243:

```

GTCCGGGTCC CACCACCCGG CCGGCGCGGC CCTAGCGGCC GGGCGCACCA GCCCCTTTTC 60
TTGACTCGTT CAAGAAAAGG GCCTTCTGTT TGGTCGGCCA TGTTGGCATG ATCGTGACCC 120
ATGGGCAACA TCGACGTGCA CATCTCGGCC AAGGTCTAGC TCCATGCGAA TCGCCGCGCG 180
GGTGGTGAGC ATCGGTCTAG CCGTCATAGC AGGTTTCGCG GTACCTGTTG CCGACGCACA 240
CCCGTCGGAG CCGGGGGTTG TGTCTACGC GGTGCTCGGA AAGGGGTGCG TCGGCAACAT 300
CGTCGGCGCC CCAATGGGGT GGGAGGCGGT GTTCACCAAG CCGTTCCAGG CTTTTGGGT 360
CGAACTACCG GCGTGCACCA ACTGGGTGGA CATCGGGCTG CCGAGGTGT ACCACGATCC 420
CGAC

```

(2) INFORMATION FOR SEQ ID NO:244:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 317 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:244:

```

CCGATGGGCG CCGCGGGTAA CAGGCGCAAT TTGGAACGGT TTGCGAAGCG CAAAGATCGT 90
TTGATCTGCG CCGCAATTGA TTGAGCCCGG GCGGATTTCG TCACCGAGGG CCACCGTCTA 120
AGGGCGGATG CGATCCTACT GCGGCGTACC GACCGGCTGC CTTTCCCGGA GCGGCGGGAT 180
TGGGATTTCG TCGAGTCCCA TTGCGGCACG ACCGTCACCG CCGACACGGT GCGCATCGAC 240
GTGATCGCGG ACGATATCGT TCGGAACTG GCGGCGCGGT CCAAACTCAG CGAATCGCTG 300
CGGCTCTAGC ATTGCTC

```

(2) INFORMATION FOR SEQ ID NO:245:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:245:

TGGCGTATGC	GCTTCGCAGC	CGGTGCCGCG	TCAACGCGCC	GGAGGCAATC	GCTTCGCTGC	60
CGAGGAATGG	TTGATCAGC	ATCGCAGTGT	GCCGTCTGTC	ACCGACACCG	CCGTCCAACG	120
TGAACTGAGG	GCGGAAAATC	GGCCGAAATC	TGCGCCTCAG	TTACAGCTCG	GCCCTTAACG	180
GTTCCTGGAAG	TTGGGTGCGC	GCTTCTCGGC	GAACGCGCGC	GGGCCTTCCT	TGGCGTCGTC	240
GGACAGGAAG	ACCTTGATGC	CGATCTGGGT	GTCCATCTTG	AACGCCTCGT	TTTCGGGCAT	300
GCACTCGGTC	TCGCGGATGG	ACCGCAAGAT	GGCCTGCACG	GCCAGGGGTC	CGTTAGCCGA	360
GATGGCGTCG	GCAAGTTCTA	GAACCTTGGT	CAACGCCTGG	CCGTGCGGCA	CACGTGGCCG	420
AT						422

(2) INFORMATION FOR SEQ ID NO:246:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 426 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:246:

GCCTGCGGCT	GAACACCAGC	CGCGGGGTGC	CAGATCTCCC	GGACTCGGTA	GTGCGGCGGG	60
TGGCGTCGTT	GCTCTCCTGA	CGGGGCGCGG	CGACCATAAG	GTGCTAATG	CCCAGGTAGC	120
GGCCGAGGTG	CATGGAGTCC	ATGATGATGC	GACTCTCCAG	CTGCGCGACC	GGGAGCTTGG	180
CATCGGGGCT	GATCAGCCAG	GACCGGTAGG	ACAAGTCGAT	CGAATGCATA	GTGGCCTCCA	240
GAGTGGGCGT	GGCAGTCCCG	GCCTGCTCCA	CGGCAAATGC	CTTGATTCTT	AGCTCCGCGT	300
AGTATTTCCG	CATCSCCTGC	GGGATGAATG	GGAACCGCAG	GATGGCGACA	AACGGGTCTG	360
ACCTGAGGTT	TGCGGCTTTG	GGCAGCTGG	TCCACAGCCG	GTACTCGGCA	TAAATGCTGG	420
CCCCGA						426

(2) INFORMATION FOR SEQ ID NO:247:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 327 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:247:

AGACCGGCGA	GGGTGTGGTC	GCTGCGGCGG	GGATTCTGGA	TAATCTGCGC	TGGGTGAGC	60
CGCCGATCAA	CTAGTGAGGT	GCACGCTTAC	GGTTTGGGAT	ACCGACAGCT	AAAAAGTTTA	120
TCAAAGAAAC	GAAGAAGGTT	TCATGAGTAA	CTTTTGGGCG	TACCGCGCGC	AGCTCGGCGG	180

(2) INFORMATION FOR SEQ ID NO:248:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 123 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:248:

```

Asp His Gly Gly Pro Ala Thr Asn Pro Gly Ser Gly Ser Arg Gly Gly
 1             5             10             15
Ala Gly Gly Ser Gly Gly Asn Gly Gly Ala Gly Gly Asn Ala Thr Gly
 20             25             30
Ser Gly Gly Lys Gly Glu Ala Gly Gly Asn Gly Gly Asp Gly Ser Phe
 35             40             45
Gly Ala Thr Ser Gly Pro Ala Ser Ile Gly Val Thr Gly Ala Pro Gly
 50             55             60
Gly Asn Gly Gly Lys Gly Gly Ala Gly Gly Ser Asn Pro Asn Gly Ser
 65             70             75             80
Gly Gly Asp Gly Gly Lys Gly Gly Asn Gly Gly Ala Gly Gly Asn Gly
 85             90             95
Gly Ser Ile Gly Ala Asn Ser Gly Ile Val Gly Gly Ser Gly Gly Ala
100             105             110
Gly Gly Ala Gly Gly Ala Gly Gly Asn Gly Ser
115             120

```

(2) INFORMATION FOR SEQ ID NO:249:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 104 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:249:

```

Met Ala Ala Ala Gly Thr Thr Ala Asn Val Glu Arg Phe Pro Asn Pro
 1             5             10             15
Asn Asp Pro Leu His Leu Ala Ser Ile Asp Phe Ser Pro Ala Asp Phe
 20             25             30
Val Thr Glu Gly His Arg Leu Arg Ala Asp Ala Ile Leu Leu Arg Arg
 35             40             45
Thr Asp Arg Leu Pro Phe Ala Glu Pro Pro Asp Trp Asp Leu Val Glu
 50             55             60
Ser Glu Leu Arg Phe Thr Val Thr Ala Asp Thr Val Arg Ile Asp Val
 65             70

```

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:250:

Ala Tyr Ala Leu Arg Ser Arg Cys Arg Val Asn Ala Pro Glu Ala Ile
1 5 10 15
Ala Ser Leu Pro Arg Asn Gly Ser Ile Thr Ile Ala Val Cys Arg Arg
20 25 30
Ala Pro Thr Pro Pro Ser Asn Val Asn
35 40

(2) INFORMATION FOR SEQ ID NO:251:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:251:

Val Pro Leu Asn Thr Ser Pro Arg Leu Pro Asp Leu Pro Asp Ser Val
1 5 10 15
Val Pro Pro Val Ala Ser Leu Leu Ser
20 25

21 INFORMATION FOR SEQ ID NO:252:

1. SEQUENCE CHARACTERISTICS

- A LENGTH: 61 amino acids
 B TYPE: imino acid
 C STRANDEDNESS: single
 D TOPOLOGY: linear

11. MOLECULE TYPE: protein

(XX) SEQUENCE DESCRIPTION. SEQ ID NO:252.

Met Ser Thr Val Ala Ala Tyr Ala Ala Met Ser Ala Thr Glu Pro Leu

Gln Thr Glu Trp Gly Gln Pro Asn Leu Pro Val Val Pro
 50 55 60

(2) INFORMATION FOR SEQ ID NO:253:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 213 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:

GCTTGGAGCC CTGGAGCGAC GGTGTGGGTC TGGGGGTGGA TTCGTTCTCG GCGAAAGTCA 60
 ACTAAAGACC ACGTTGACAC CCAACCGGCG GCCCGGCATG GCGCGTCGCG GCGTAGAAGC 120
 TTTGACCGCG GCGCGAAACG TTCGCTGCTG CGGCCCATGC AGATCGCACA CGCTTGCTTG 180
 AACATCGGGT GGAGCCCGTG CTAACGCCAG SCT 213

(2) INFORMATION FOR SEQ ID NO:254:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 367 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:254:

CCGAGCTGCT GTTCGGCGCC GCGCGTGGCG GCGCGCGCGG TGGGGGCGGC ACCGACGGCG 60
 GCGCCCGTGC TACCGCGCGG ACCCGCGGAT ACCCGCGAGT CGCGCGCGAC GCGCGATGGC 120
 TGGCACCCCG CGGGGCGCGC GGGGCGCGCG GCGAAGGCGG GCGAGGTGGT CCGCGCAGCG 180
 ATGGTGGCGC GTTGGGTGGT ACCGCGCGGA TGGCGGTAC CGCGCGCGCC GGTGGCGCGC 240
 GCGGTGCGCG CACACTGCTG CTGGGCGCTG GCGGACAGGG CGGCTTCGGC GCGCGCGCGC 300
 GACGAGGCGG CACCGCGCGG GCGCGCGCGA GATGGCGTTC TGGGGGTGT CAGTGGCACT 360
 GTGGTA 367

(2) INFORMATION FOR SEQ ID NO:255:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 420 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

```

GTCCGCTCTA TGGGACGGGT GCGACGCGGC ATTGTCCAAC CGCCTGGGCT ACCCCGAAGT      240
CCGCGCGCGCA CTCGCTGCAA CGGGCGCGCA TCACGACCTA ACCGAATCCG AGCTCGCGGA      300
CGCCGAGCGT GACTGGGAGG ACTTCTGGGC CGCACCCGCC CAGTCGAACT CACCGCGACG      360
GTTGAACAGC ACGCCGGGCA CCTCGCCCGA ACACATGCCT TACGCGGAGC CGACACCGTT      420

```

(2) INFORMATION FOR SEQ ID NO:256:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 299 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:256:

```

CTTTTGTGGG TGGCATCGGC GGTACCGGCG GAACCGGGCG CAACCGCGGT ATGCTCGCGG      60
GGCGCGCGCG GCGCGCGCGT GCGCGCGGGT TTAGCTTCAG CACTGCGCGT GGGGCTGGCG      120
GGCGCGCGCG GCGCGGTGGG CTGTTACCCA CCGCGGGTGT CGGCGGCGCG GGTGGGCAGG      180
GTCACACGGG CGGCGCGCGC GCGCGCGGCG GGGCGGCGCG GTTGTTTGGT GCGCGCGCGA      240
TGGCGCGGGG GGGCGGATTC GGGGATCAGG GAATGCTCGG CACCGGCGCG GCGCGCGGG      299

```

(2) INFORMATION FOR SEQ ID NO:257:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:257:

```

Leu Glu Pro Trp Ser Asp Gly Val Gly Leu Gly Val Asp Ser Phe Ser
 1             5             10             15
Ala Lys Val Asn
 20

```

(2) INFORMATION FOR SEQ ID NO:258:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 121 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:258:


```

      20      25      30
Val Gly Gly Asp Gly Gly Trp Leu Ala Pro Gly Gly Ala Gly Gly Ala
      35      40      45
Gly Gly Gln Gly Gly Ala Gly Gly Ala Arg Ser Asp Gly Gly Ala Leu
      50      55      60
Gly Gly Thr Gly Gly Thr Gly Gly Thr Gly Gly Ala Gly Gly Ala Gly
      65      70      75      80
Gly Arg Gly Thr Leu Leu Gly Ala Gly Gly Gln Gly Gly Leu Gly
      85      90      95
Gly Ala Gly Gly Gln Gly Gly Thr Gly Gly Gly Arg Arg Arg Trp Arg
      100      105      110
Ser Gly Gly Cys Gln Trp His Trp Trp
      115      120

```

(2) INFORMATION FOR SEQ ID NO:259:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION. SEQ ID NO:259:

```

Gly Val Ile Gly Lys Ala Thr Ala Gln Arg Pro Val Ala Ala Gly Arg
 1           5           10           15
Pro Arg Pro Arg Pro Gln Arg Pro Val Ser Asp Arg Val Ser Asp Gln
      20           25           30
Arg Arg

```

2. INFORMATION FOR SEQ ID NO:260:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 99 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION. SEQ ID NO:260:

```

Leu Val Gly Gly Ile Gly Gly Thr Gly Gly Thr Gly Gly Asn Ala Gly
 1           5           10           15
Met Leu Ala Gly Ala Ala Gly Ala Gly Gly Ala Gly Gly Phe Ser Phe
      20           25           30
Ser Thr Ala Gly Gly Ala Gly Gly Ala Gly Gly Ala Gly Gly Leu Phe
      35           40           45

```

Gly Gly Ala Gly Gly Phe Gly Asp His Gly Thr Leu Gly Thr Gly Gly
 85 90 95
 Ala Gly Gly

(2) INFORMATION FOR SEQ ID NO:261:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 282 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:261:

TCCTGTTCCG	CGCCGCGCGG	GTGGGCGGTG	TTGGCGGTGA	CGGTGTGGCA	TTCTGGGCA	60
CGGCCCCCGG	CGGGCCCCGT	GGTGCCGCGG	GGGCGCGGTG	GCTGTTGAGC	GTCCGGTGGGG	120
CGGGCGGCGG	CGGCGGAATC	GGATTGGTGG	GGAACAGCGG	TGCCGGGGGG	TCCGGCGGGT	180
CGGCGCTGCT	CTGGGGCGAC	GGCGGTGCGG	GCGGCGCGGG	TGGGCTCGGG	TCCACTACCG	240
GGGGTSCCGG	CGGGGCGGGG	GGCAACGCCA	GCTGCTGGT	AA		282

(2) INFORMATION FOR SEQ ID NO:262:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 415 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:262:

CGGCACGAGC	CGTGCTACTG	GTCAACTGAT	GGCTGATTG	TGACCTTCCG	GGCGCCGGAT	60
CAGTGCTTCT	CAGGACCGAC	GTAAATATCG	AAACCAATC	CGGCGCCCGA	GGCGAGGATG	120
AATGCCACAC	CGGCGGCGAT	CAGCCACGGG	AGCCACAACG	CGATGCCCGAC	CGCTGCCACC	180
GAGCCGGACA	ACGCGACCAT	GATCGGCCAC	CAGCTATGCG	GACTGAAGAA	TCCAAATTCT	240
GCTGCGCCGT	GGCTGATTTC	AGGCGCTTCT	TAGTCCCTCG	GCCGGGAATC	TAACCGGCGG	300
AGACAAACG	GGAAGAAGGT	GCGGACGATC	AACGCCATGC	CGCGGCTGAG	CGCCAAACGCA	360
ATGGTCCGAG	CCCACTGAGC	AGCAAGGCTG	GGAACATCG	AGGTCAACAC	GGCGT	415

(2) INFORMATION FOR SEQ ID NO:263:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 373 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

TCACCGCGTG AACGGTTCGT AACACTGATA CGTATGCTTG TCAGCGAGCA GATCAAGTCC	60
AGTCCGACCA ATGCCAGGAG ATCATCGGCT AGGCTCACGG TTTCGCCTGG GACGAGACGG	120
TATTGAGTTC TGGCCTTGA CGGTCCGTGG CGTGGTGGGA AGTCTGACGC GGCATCAGAA	180
CGGTGTGCAA TACCAGTCTT TGGGGGATAT GGCCTATTTG GTGTCGTGG GCCGCTCCAC	240
CGGATCCCTT TTCGAACGTT GC3CAAGCGC GGTCCAGTTA CGGCCTGTTC ACTGCGCGCT	300
GGCGTAGCTG CGCGGCTCG ATCGGTTTGA ACGTCATCGC AATTCGCCGA ATGGGTGAGT	360
ACCTGACGCT CCT	373

(2) INFORMATION FOR SEQ ID NO:264:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 423 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:264:

CCAAACCGGA CAGGCCGGCA GCGACGGTCG GAAGTTGCAC CAUGGTGCGC GCTCCATGTA	60
GCGAACCGGT GACCACGGCG TAGACAGCAG ATCCGTGGAT CGGCGGTTCC GTGTCGTCCG	120
GGCCGAGTAC CCGCGGGCCG AACCGCAGCG ACCAAAGCAA CGCGATCGAT ACGGGGATCG	180
CCACTCGTGC CCAATTGAG CTCCCTCGAC AAGCTTGGCG CCGCACTCGA ACCCGGGTGA	240
ATGATTGAGT TTAACCGCT TAGCAATAAC TAGCATAACC CCGTGGGGCC TCTAAACGGG	300
TCTTGAGGGG TTTTGTGCTG AAAGGAGGAA CTATATCCGG ATAACCTGGC GTAGTAGCGA	360
AGAGGCCCGC ACCGATCGCC CTTCCCAACA GTTGCGGAGC GTGAATGGCG AATGGACGCG	420
CCC	423

(2) INFORMATION FOR SEQ ID NO:265

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 404 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:265

ACTGCGCAGC CGGTGCGCA ATCATTCAGC TTTCGTTTAC ATGAGTGCAC CATCGCGGCA	60
ACTGACCTGC ACCGAGTCAG CTCTTTCGCG AGGATTTTTC TGAACCGGG TATGCACTGC	120
CGGTCCCGAC GTGCTGACAT CCGCCAGGCT AGGTGCGGTC GCATGCTGCT CGCCGTCCAG	180
CGAATGTGCG ACGTCGCGCA CCGCCCGATG CGACCGGCTG ACCAAGCGCT GTAACACAGC	240
CAGCACCGGC CTGCGGAGGC TTTCGGGCGT CATGCTGCTG AGGCGACCGT CGAGCAGCTC	300
CAGTACGAGC GTGAGCTCAC CGGTGCTGCG GTGCGCGGCG AGGCTCAGCG GAAAGTGCGA	360
GAAGTCTCT AGGCGTACCG GAGCGAACCT CAGCGCTTTT GCGA	404

(2) INFORMATION FOR SEQ ID NO:266

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:266:

```

GTCCTGGTCC CAGGCTGTTT TTCGAACCCG CTGGCTAACT TCGCACCCCG GTATCCGCCC      60
ACCATCGAAC CCGCCCAACC GCGCGTGTCG CCGCCTACTT CGCAAGACCC GGCCGGTGCA      120
GTGCGACCAC TGAGCGGCCA CCCCCGGGCG GCACTATTCTG ACAACGGCAC CCGCCAATTG      180
GTGGCTCTGC GCCCGGGGCG CGATTGGGCG GCACCCGCCA GCATCATGGT CTTGATGAC      240
ATGCACGTTG CACCGCGCGT CATTITCTG CCGGGCCCCG CAGCCGCGTT GACCAGCGAC      300
GACCACGGCA CGGCCTTCCT TGCCGCCCGC GCGCGCTACT TCGTGGCCGA CCTGTCTCTC      360
GGTCACACCG CACGAGTGAA TGTCGCTGAC GCAGCGCACA CCGATTTCAC CGCGATCGCC      420
C

```

(2) INFORMATION FOR SEQ ID NO:267:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 426 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:267:

```

ATGCATATCA CGCTCAACCG CATCTGCCCT GCGATCTTCG GGGCCGGCGG CAGTGAACCTA      60
GACGAGCTGC GCGCGCTCAT TCGCGCGTGG GTCACGCTGG GCTCGCGCCT GGCGGGCGCTA      120
CCGAAACCCA AACCGGACTA TGGCGCGCTT AGCCCGTGGG GCGCGCTGGC CGAGTGGCGG      180
CCCGAGTACG ACACTGTCTT CGACGAGCTG ATCGAAGCCG AGCGGGCCGA CCGGAACCTC      240
CCCGATCGGA CCGACGTTTT GCGGTTGATG CTGCGCAGCA CTTACGACGA CGGTTCCATC      300
ATCTCGCGCA AGGACATTGG CGACGAACTG CTCACGCTGC TTGCGGCCCG GCACGAAACC      360
ACCGCGCGGA CATGGGCTGG CGCTTCGAAC CGCTCAACCG GCACCGCGAG GTGCTCGCGG      420
CTCTGG

```

(2) INFORMATION FOR SEQ ID NO:268:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 426 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:268:

```

GTCTGCTGCG CAGGCTGTTT TTCGAACCCG CTGGCTAACT TCGCACCCCG GTATCCGCCC      60
ACCATCGAAC CCGCCCAACC GCGCGTGTCG CCGCCTACTT CGCAAGACCC GGCCGGTGCA      120
GTGCGACCAC TGAGCGGCCA CCCCCGGGCG GCACTATTCTG ACAACGGCAC CCGCCAATTG

```

CGCCGCTCCG ACGGCAAGCT GGTGCTGGGG AGCGCAGATG GCGCCGTCTA CACGCTTGCC 480
AAGAACCCGC AGTTGACCGG CGTCGGCGGC GCCACCGTAG CC 522

(2) INFORMATION FOR SEQ ID NO:269

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 739 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:269:

GCTGGGGCGC	ACCGCCGTCC	GGCGGCCCA	GCCCTGGGC	CCAGACCCG	CGCAAAACCA	60
ACCGGTGGC	CTTACTGGC	GGGUCGCGC	CCGTCTGCT	CGTCTCGTG	TTGGGCGCCA	120
TGGGCATCTG	GATCGCCATG	CGGCCCAAGC	CGGTACAGCC	GCCTCAGCCG	GTTGCGGAGG	180
AGGCGCTTAG	CGCCCTACTG	GTGAAGTCCT	CAGAAGTCAA	CGCCGTGATG	GGCTCGTCTG	240
CCATGCAAGC	GGGCAAACCG	ATCAGATCGA	TGGACTCTTC	GCCGGTGACG	GTGTCCCTGC	300
CGGACTGCCA	GGGCGCGCTG	TATACCAGCC	AGGATCCGGT	GTATGCCGGC	ACCGGCTACA	360
CGGACATCAA	CGGCTTGATT	TCATCCGAGC	CGGGCGACAA	CTACGAACAT	TGGGTGAACC	420
AAGCGTCTGT	CGCCTTTCCG	ACCGCCGACA	AAGCCCGCGC	GTTCGTGCAG	ACTTCGGCCG	480
ACAAATGGAA	GAAGTGGCA	GGCAAGACGG	TCACCGTCAC	GAATAAGGCC	AAGACCTACC	540
GGTGGACGTT	TGCCGACGTC	AAAGGCAGCC	CGCCGACGAT	CACGGTGATA	GACACCCAAG	600
AAGGCGCTGA	GGGCTGGGAA	TGCCAACGCG	CGATGAGCGT	GGCCAACAAT	GTGGTTGTCT	660
ACGTCAACGC	ATGCGGGTAC	CAGATCACCA	ATCAAGCAGG	CCAGATCCGC	GCCAAGATCT	720
GTGACAAAG	TCAACAAGG					739

(2) INFORMATION FOR SEQ ID NO:270:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 69 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:270:

AGAGTCTCTC GAGGCGCGCA GCGCGCGC TGAATCTCTC AGGCGCGC TGAACGCTCT 60
GGTATCTC

(2) INFORMATION FOR SEQ ID NO:271:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 523 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

ACTGCACCCG	GCAGGCGGCA	CCAACGGATC	GGGTCAACTA	GCACTGCCCG	TGGAGGCGCC	60
CCCGCGGTCT	GTGCCCTTCC	ACGGGGAACC	TTTGGGCAGC	GCGGCTCCAG	AAGGGTTGGA	120
GGGAGAGTTC	GACGACCGTA	TCGACGAGCG	GTTCCCGGTC	TTGAGCTCGG	CCAGTCTCGC	180
CGAAGCGCTG	CCGGGTCCGC	TGACCCCGAT	GACGCTGGAT	GTCCAGTTGA	GTGGACTGCG	240
CGCGGCCGGT	CGGGCGATGG	GTCGGGTACT	GGCGCTTGGC	GGTGTCTGTT	CCGATGAGTG	300
GGAGAGAAGA	GCCATCGCGG	TGTTTCGGTCA	CCGCCCCTAT	ATCGGAGTGT	CGGCCAATAT	360
TGTGGCCGCC	GCCCAACTGC	CGGGGTGGGA	CGCGCAGGCC	GTAACCCGGC	GGGCACTGGG	420
CGAGCAACCG	CAGGTCAC TG	AGCTGCTTCC	GTTTGGTCTGA	CCGCAACTTG	CGGGCGGACC	480
GCTCGGCTCG	GTGCGGAAGG	TGGTCTGAC	GGCACGGTCTG	CTG		523

(2) INFORMATION FOR SEQ ID NO:272:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 224 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:272:

GTGTGGGTGT	CGTCGGGGTA	GGAGCGACTT	CCCGGGCCGG	CGCGGGCGCC	GGAGCGGGCT	60
GTGCGGAAC	CGGTCCCGGC	GTCGCGCGCG	GGGCGACCAA	AGGCCGGATC	GATTCGGCCA	120
CGCGCTTGGC	CGCGCCCTTG	TTCACCGGCT	TGTTGGCGGT	CCCGAGCCAT	ACCACAAACC	180
AACGCTGAAG	GGCCCCCGCG	TCCGCTGGCT	TCCCGCGGGG	CGAC		224

(2) INFORMATION FOR SEQ ID NO:273:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 521 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:273:

TGAAGTGAAT	TCGGGCTGAG	ATGCGCGGCG	TTTCTTCTTT	ATAGCTGGCG	TCGAGGCGCA	60
TGAAGTGTCT	TTGGGCTAG	CGGGGCTTGG	TCTCTTCTTT	GTCCAAAGCC	TGCAGGCGCT	120
TTTACTGGCG	TTGGTGAAG	TCGAGCTAG	TTTCTTCTTT	ATTCGAGG	CGATCGGCT	180
TCTGCAACGC	CAGATCGGCG	GTGCTGATCG	TCGCGGCGAG	TAAGGAGGTC	TAGAGCACTT	240
TCTGCAATAG	GTGCTTCTTT	TCGATCAGCT	TCGCTTCTTT	TAAGGAGGTC	TCGAGGCGCT	300
TCTGCGTCAG	GCGGACATCG	AGGCGGCGCG	TGAAGAGGTC	TAAGGAGGTC	TCGAGGCGCT	360
TCTGCGTGGC	CAGCAACATG	AGGCTGCGAG	TCTTCTTCTTT	TAAGGAGGTC	TCGAGGCGCT	420
TCTGCGCACT	CTGATCTGTC	GACCAAAATG	TCGGAATCTT	TCGCGGTCGG	CTGCGGAGCG	480
TCTGCAATAG	GCGGAGGTCG	TGCGGAGGTC	AAGGAGGTCG			521

(2) INFORMATION FOR SEQ ID NO:274:

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:274:

```
CTCAGGCTC ATTGCTCGA ACAAAGCCAC CCGCCCGTAC AGCGGACGCC CCCATTCGTT      60
GTCGTGATAG TCGCGGTACA GCTGGGCATC GGGCCCTGGA CGAACCTCCG CCCAGGGGCA      120
GCGAACCAGC CCGTCGCCGC TCACGCGGGG TCAGAACGGT AGTGCACGAC AGTCTCGCCG      180
CCGGAAGGGT TTGACGCGTC AGACTCGGCC TCGGCGTCTT CCGACGAGGC GTGGATCGCC      240
CCGAGCTGAG AGCGTAGCGC CTCGAGCTCA CCGCCGAGCC GTTCCAGCAC CCAGTCCACC      300
TCGCTGGTCT TGTTCGCCG CAGCACCTGC GTGAACCTGA CCGCGTCCGAC ATCGGCSCCG      360
GTGACCCCGA ACSCCGGCAG CSTCGTCGCC GTCTCGCCC GCGGCAGGGG CCGCAACTGC      420
TCGCCA                                     426
```

(2) INFORMATION FOR SEQ ID NO:275:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 219 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:275:

```
GCGGACACGG CCGACAAAGC GCAATCGGCC TCGGCGGCGG CCGCGGCGGC GACGGGGGCC      60
AGGGCGGCGC CCGCCCGGGA TTGTGGGATA CTGGCGGCGC CCGCGGACAC GCGGGGGCAA      120
CGCGGTGATA CCGGGGCGCC ACCGCTGCCC GGTGAGGCAG GCATGGGCGC CCGGGGTGGC      180
CGCGGTGGG GTATCGGCAA CGCGGCGGCG CCGGGCGGAC                                     219
```

(2) INFORMATION FOR SEQ ID NO:276:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 571 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:276:

```
AAGATCATCG GCGCGGCTCG TTAGCATCGG TCGCTCTCTG ATGCTGCGCG GCGCGGATCA      60
TGGAGGCTCG GCGCTTCTAC CCACTGCTCG AAGGCTGAGC ACCACAGTCG GGTCTCTCGG      120
ATCGTTTTCG ACCTTGGCGC CCAGACGCTG TATATGCGCG TTGACGAGCG TGGTATCGCG      180
TGGGTGCGCG TAAGGCGATA CCGCTTCTAG TATAGATCGA CAGTAAAGCA CCGGGGCGCG      240
CTTCCGCGCG AATTTGAGCA ACAGCTCGAA TTGAGGCTGT CTCAGGCTGT CTCAGGCTGT      300
ATTCGAGCTG AGCTTCTCTG CCGCTTCTCT TCGCTCTCTG ATGCTGCGCG GCGCGGATCA      360
```

(2) INFORMATION FOR SEQ ID NO:277:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 93 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:277:

```

Leu Phe Gly Ala Gly Gly Val Gly Gly Val Gly Gly Asp Gly Val Ala
 1             5             10             15
Phe Leu Gly Thr Ala Pro Gly Gly Pro Gly Gly Ala Gly Gly Ala Gly
 20             25             30
Gly Leu Phe Ser Val Gly Gly Ala Gly Gly Ala Gly Gly Ile Gly Leu
 35             40             45
Val Gly Asn Ser Gly Ala Gly Gly Ser Gly Gly Ser Ala Leu Leu Trp
 50             55             60
Gly Asp Gly Gly Ala Gly Gly Ala Gly Gly Val Gly Ser Thr Thr Gly
 65             70             75             80
Gly Ala Gly Gly Ala Gly Gly Asn Ala Ser Leu Leu Val
 85             90

```

(2) INFORMATION FOR SEQ ID NO:278:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:278:

```

Met Pro Pro Val Ser Ala Asn Ala Met Val Glu Val His Ser Thr Pro
 1             5             10             15
Met Val Ala Asn Ile Glu Val Asn Thr Pro
 20             25

```

(2) INFORMATION FOR SEQ ID NO:279:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

Trp Pro Ala Gly Arg Pro Met His Pro Ala Pro Gly Thr Ser Ala Asp
1 5 10 15
His Pro Pro Asn
20

Val	Leu	Val	Ala	Gly	Asn	Ser	Ser	Asn	Pro	Leu	Ala	Asn	Phe	Ala	Pro
1				5				10						15	
Gly	Tyr	Pro	Pro	Thr	Ile	Gln	Pro	Ala	Gln	Pro	Ala	Val	Ser	Pro	Pro
		20					25						30		
Thr	Ser	Gln	Asp	Pro	Ala	Gly	Ala	Val	Arg	Pro	Leu	Ser	Gly	His	Pro
		35					40					45			
Arg	Ala	Ala	Leu	Pro	Asp	Asn	Ile	Thr	Arg	Ile	Leu	Val	Ala	Leu	Arg
		50				55					60				
Leu	Gly	Ala	Asn	Ser	Ala	Ala	Pro	Ala	Ser	Ile	Met	Val	Phe	Asp	Asp
65											70				
Met	His	Val	Ala	Pro	Arg	Val	Ile	Phe	Leu	Pro	Gly	Pro	Ala	Ala	Ala
				75					80						
Leu	Thr	Ser	Asp	Asp	His	Gly	Thr	Ala	Phe	Leu	Ala	Ala	Arg	Gly	Gly
		100						105					110		
Tyr	Phe	Val	Ala	Asn	Leu	Ser	Ser	Gly	His	Thr	Ala	Arg	Val	Asn	Val
		115						120					125		
Ala	Asn	Ala	Ala	His	Thr	Ala	Thr	Pro	Ala	Ala	Ala	Ala	Ala	Ala	Ala

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 142 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:282:

```

Met His Ile Thr Leu Asn Ala Ile Leu Arg Ala Ile Phe Gly Ala Gly
 1           5           10           15
Gly Ser Glu Leu Asp Glu Leu Arg Arg Leu Ile Pro Pro Trp Val Thr
 20           25           30
Leu Gly Ser Arg Leu Ala Ala Leu Pro Lys Pro Lys Arg Asp Tyr Gly
 35           40           45
Arg Leu Ser Pro Trp Gly Arg Leu Ala Glu Trp Arg Arg Gln Tyr Asp
 50           55           60
Thr Val Ile Asp Glu Leu Ile Glu Ala Glu Arg Ala Asp Pro Asn Phe
 65           70           75           80
Ala Asp Arg Thr Asp Val Leu Ala Leu Met Leu Arg Ser Thr Tyr Asp
 85           90           95
Asp Gly Ser Ile Met Ser Arg Lys Asp Ile Gly Asp Glu Leu Leu Thr
100          105          110
Leu Leu Ala Ala Gly His Glu Thr Thr Ala Ala Thr Trp Ala Gly Arg
115          120          125
Ser Asn Gly Ser Thr Gly Thr Pro Thr Cys Ser Arg Leu Trp
130          135          140

```

(2) INFORMATION FOR SEQ ID NO:283:

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 163 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

xi. SEQUENCE DESCRIPTION: SEQ ID NO:283:

```

Val Leu Val Ala Gly Val Ser Ser Asn Ser Leu Ala Asn Pro Ala Ser
 1           5           10           15
His Tyr Pro Pro Thr Ala Glu Pro Ala Glu Ser Ala Val Ser Pro Pro
 20           25           30
Thr Ser Gln Asp Pro Ala Gly Ala Val Arg Pro Leu Ser Gly His Pro
 35           40           45
Arg Ala Ala Leu Phe Asp Asn Gly Thr Arg Gln Leu Val Ala Leu Arg
 50           55           60
Pro Gly Ala Arg Ser Ala Ala Thr Ala Ser Leu Met Val Phe Asn Asn
 65           70           75           80

```

Tyr Phe Val Ala Asp Leu Ser Ser Gly His Thr Ala Arg Val Asn Val
 115 120 125
 Ala Asp Ala Ala His Thr Asp Phe Thr Ala Ile Ala Arg Arg Ser Asp
 130 135 140
 Gly Lys Leu Val Leu Gly Ser Ala Asp Gly Ala Val Tyr Thr Leu Ala
 145 150 155 160
 Lys Asn Pro

(2) INFORMATION FOR SEQ ID NO:284:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 240 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:284:

Trp Gly Ala Pro Pro Ser Gly Gly Pro Ser Pro Trp Ala Gln Thr Pro
 1 5 10 15
 Arg Lys Thr Asn Pro Trp Pro Leu Val Ala Gly Ala Ala Val Val
 20 25 30
 Leu Val Leu Val Leu Gly Ala Ile Gly Ile Trp Ile Ala Ile Arg Pro
 35 40 45
 Lys Pro Val Gln Pro Pro Gln Pro Val Ala Glu Glu Arg Leu Ser Ala
 50 55 60
 Leu Leu Leu Asn Ser Ser Glu Val Asn Ala Val Met Gly Ser Ser Ser
 65 70 75 80
 Met Gln Pro Gly Lys Pro Ile Thr Ser Met Asp Ser Ser Pro Val Thr
 85 90 95
 Val Ser Leu Pro Asn Lys Gln Gly Ala Leu Tyr Thr Ser Gln Asp Pro
 100 105 110
 Val Tyr Ala Gly Thr Gly Tyr Thr Ala Ile Asn Gly Leu Ile Ser Ser
 115 120 125
 Gln Pro Gly Asn Asn Tyr Glu His Thr Val Asn Gln Ala Val Val Ala
 130 135 140
 Leu Tyr Thr Ala Asp Lys Ala Arg Ala Thr Val Gln Thr Ser Ala Asn
 145 150 155 160
 Lys Thr Lys Asn Lys Ala Gly Lys Thr Val Thr Val Thr Asn Lys Ala
 165 170 175
 Lys Thr Tyr Arg Trp Thr Phe Ala Asn Val Lys Gly Ser Pro Pro Thr
 180 185 190
 Ile Thr Val Ile Asn Thr Gln Gln His Ala Gln Gly Tyr Gln Cys Gln
 195 200 205
 Arg Ala Met Ser Val Ala Asn Asn Val Val Val Asn Val Asn Ala Cys
 210 215 220
 His Thr Gln Ile Thr Asn Gln Ala Thr Thr Thr Thr Thr Thr Thr Thr

238

- (A) LENGTH: 22 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:285:

```

Asp Val Val Glu Ala Ala Ile Ala Arg Ala Glu Ala Val Asn Pro Ala
 1             5             10             15
Leu Asn Ala Leu Ala Tyr
                20

```

(2) INFORMATION FOR SEQ ID NO:286:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 174 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:286:

```

Leu His Pro Ala Gly Ala Thr Asn Gly Ser Gly Gln Leu Ala Leu Pro
 1             5             10             15
Val Glu Ala Pro Pro Arg Ser Val Pro Ser His Gly Glu Pro Leu Gly
 20             25             30
Ser Ala Ala Pro Glu Gly Leu Glu Gly Glu Phe Asp Asp Arg Ile Asp
 35             40             45
Glu Arg Phe Pro Val Phe Ser Ser Ala Ser Leu Ala Glu Ala Leu Pro
 50             55             60
Gly Pro Leu Thr Pro Met Thr Leu Asp Val Gln Leu Ser Gly Leu Arg
 65             70             75             80
Ala Ala Gly Arg Ala Met Gly Arg Val Leu Ala Leu Gly Gly Val Val
 85             90             95
Ala Asp Glu Thr Gln Arg Arg Ala Ile Ala Val Phe Gly His Arg Pro
100             105             110
Thr Ile Gly Val Ser Ala Asn Ile Val Ala Ala Ala Gln Leu Pro Gly
115             120             125
Pro Asp Ala Gln Ala Val Thr Arg Arg Ala Leu Gly Gln Gln Pro Gln
130             135             140
Val Thr Gln Leu Leu Pro Phe Gly Arg Pro Gln Leu Ala Gly Gly Pro
145             150             155             160
Leu Gly Ser Val Ala Lys Val Val Val Thr Ala Arg Ser Leu
                165             170

```

(2) INFORMATION FOR SEQ ID NO:287:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:287:

```

Val Gly Val Val Gly Val Gly Ala Thr Ser Pro Ala Gly Ala Gly Ala
 1           5           10           15
Gly Ala Gly Ser Ala Gly Thr Gly Ala Gly Ala Gly Gly Gly Ala Thr
 20           25           30
Lys Gly Arg Ile Asp Ser Ala Ser Ala Leu Ala Ala Pro Leu Ser Thr
 35           40           45
Gly Leu Leu Ala Val Pro Ser His Thr Thr Asn Gln Arg
 50           55           60

```

(2) INFORMATION FOR SEQ ID NO:288:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 133 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:288:

```

Met Ala Asn Thr Gly Ser Leu Val Leu Leu Arg His Gly Glu Ser Asp
 1           5           10           15
Trp Asn Ala Leu Asn Leu Phe Thr Gly Trp Val Asp Val Gly Leu Thr
 20           25           30
Asp Lys Gly Gln Ala Glu Ala Val Arg Ser Gly Glu Leu Ile Ala Glu
 35           40           45
His Asp Leu Leu Pro Asp Val Leu Tyr Thr Ser Leu Leu Arg Arg Ala
 50           55           60
Ile Thr Thr Ala His Leu Ala Leu Asp Ser Ala Asp Arg Leu Trp Ile
 65           70           75           80
Pro Val Arg Arg Ser Trp Arg Leu Asn Glu Arg His Tyr Gly Ala Leu
 85           90           95
Gln Gly Leu Asp Lys Ala Glu Thr Lys Ala Arg Tyr Gly Glu Glu Glu
100           105           110
Glu Met Ala Trp Arg Arg Ser Tyr Asn Thr Pro Pro Arg Pro Ile Glu
115           120           125
Arg Gly Ser Gln Phe
130

```

(2) INFORMATION FOR SEQ ID NO:289:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 133 amino acids

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:289:

```

Pro Gly Ser Phe Ala Arg Thr Lys Pro Pro Gly Arg Thr Ala Asp Ala
 1             5             10             15
Pro Ile Arg Cys Arg Asp Ser Arg Gly Thr Ala Gly His Arg Ala Leu
      20             25             30
Asp Glu Pro Pro Pro Arg Gly Ser Glu Pro Ala Arg Arg Arg Ser Arg
      35             40             45
Gly Val Arg Thr Val Val His Asp Ser Leu Ala Ala Arg Arg Val
      50             55             60

```

(2) INFORMATION FOR SEQ ID NO:290:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 72 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:290:

```

Gly His Gly Gly Gln Ser Ala Ile Gly Leu Gly Gly Gly Ala Gly Gly
 1             5             10             15
Asp Gly Gly Gln Gly Gly Ala Gly Arg Gly Leu Trp Gly Thr Gly Gly
      20             25             30
Ala Gly Gly His Gly Gly Ala Arg Arg Trp Tyr Arg Gly Pro Thr Ala
      35             40             45
Ala Arg Ser Gly Arg His Gly Arg Arg Gly Trp Arg Arg Trp Ala Asp
      50             55             60
Arg Gln Arg Arg Gly Arg Arg Arg
      65             70

```

(2) INFORMATION FOR SEQ ID NO:291:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 74 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:291:

```

Arg His Arg Arg Arg Ser Leu Ala Ser Leu Arg Ser Ala Ser Ser Pro
 1             5             10             15
Ala Arg Ile Thr His Val Arg Pro Cys Thr Pro Leu Leu Thr Arg Arg
      20             25             30

```

Pro Ile Pro Val Arg Ala Ala His His Glu
65 70

(2) INFORMATION FOR SEQ ID NO:292:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 174 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:292:

CCGCACGTAA CACCGTGAAT TGAAGGGAGC CCCTGGTCAT GGGCCGATTC TATCCGTGGG	60
CGAACGGTTA TTGACGGCCC GGAGGCCACT CCGCTGCCAC CAAGTGGTGA CTCAGCCCGT	120
TTTCACGGCA ACCAACGGCG GACACACCA TTAGATTCC ACAGCAGGGC CGCG	174

(2) INFORMATION FOR SEQ ID NO:293:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 404 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:293:

TGCGAAACCG GGTGACCTTC CTTCCCGTGG CCTAGAGAG TTTTCTCCAC TTCCCGGTGA	60
CCCTCGCCCG GCAGCTCAGC ACCGCTCAGT TCACCTCTCT AGTGGAGGTC CTCGACGGTC	120
CCCTCGCCCG GATGCGCGCG GAAGCGCTTC TCAGCGCGGT GGTGCTCTTC TTACAGCGCT	180
TGCTCAGCCG CTGGGATCTG CGCTGCGCG ACCTCGACAT TCTGCTGAC GCGGAGCAGC	240
ATCGGACCGC ACCCGCGCTG TCGGATCTGA TGACGTCCGG ACCGCGGTG CATACCGCGT	300
TGGCGGAAAT CCCTCGCGCA CAGCTCAGT CGGTGGCGGT GAGTTGGGCG GATGCTCAGC	360
TGACCTACCG GGAGCTGGAT GATTGGCGC ACCGGCTGGT TACT	404

(2) INFORMATION FOR SEQ ID NO:294:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 134 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:294:

Leu Val Glu Val Leu Asp Gly Ala Leu Gly Thr Met Ala Pro Glu Ser
35 40 45

Leu Gly Arg Arg Val Leu Ala Val Leu Gln Arg Leu Val Ser Arg Trp
50 55 60

Asp Arg Pro Leu Arg Asp Val Asp Ile Leu Leu Asp Gly Glu His Asp
65 70 75 80

Pro Thr Ala Pro Gly Leu Pro Asp Val Thr Thr Ser Ala Pro Ala Val
85 90 95

His Thr Arg Phe Ala Glu Ile Ala Ala Ala Gln Pro Asp Ser Val Ala
100 105 110

Val Ser Trp Ala Asp Gly Gln Leu Thr Tyr Arg Glu Leu Asp Ala Leu
115 120 125

Ala Asp Arg Leu Ala Thr
130

(2) INFORMATION FOR SEQ ID NO.235:

1. SEQUENCE CHARACTERISTICS:

(A) LENGTH: 526 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

EX1: SEQUENCE DESCRIPTION: SEQ ID NO:295

GTTTTCGACGG	CTACGAGTAC	CTGTTCTGGG	TGGGTTTGTGC	GGGCGCCTAC	GACGACAAGG	60
CGAAGAAGAC	CACCAAGGGC	TTGCGCGAGC	TGTTTCGCGT	CGCGGGGTG	AAATACTTGG	120
TGCTGGGGCCC	TGGGGAACCC	TGCACGCGC	ACTCGGGCGC	CCGCTCCGCG	AACGAGTTCC	180
TCTTTCAGCA	GCTGGCACAA	CAGGCGCTCG	AGACGCTGGA	CGGTTTGTTT	GAGGGTGTGG	240
AGAAGCTCGA	CGCGAAGATC	GTTGTACCTT	GCGCGACTG	GTTTCAACAC	ATCGGCAAGG	300
AAATATCGGCA	GCTGGGCGCT	AACATACCGG	TGCTGCACCA	CACCCAGCTG	CTCAATCGGT	360
TGCTGGCGCA	CAAGAGGCTG	TTGCTGTCTA	CTCGGGTTTC	TCAGGACATC	ACCTACGACG	420
ACCTGTGCTA	GCTGGGTGGG	TACAACAAAG	TCTACGAGGC	AACACGGGAG	CTGATCGGTG	480
CGCTGGGGGTC	CACCTGAGCC	TAGATGCTCC	TCCATGCGCA	TCGCGAG		520

2) INFORMATION FOR SEC ID NO: 094

SECRET

A LENGTH: 48" base pairs

2 THE. 10-11-1910

STRANDEDNESS

TOPOLOGY Linear

11	MOLECULE TYPE	CDNA
12	1	1
13	2	2
14	3	3
15	4	4
16	5	5
17	6	6
18	7	7
19	8	8
20	9	9
21	10	10
22	11	11
23	12	12
24	13	13
25	14	14
26	15	15
27	16	16
28	17	17
29	18	18
30	19	19
31	20	20
32	21	21
33	22	22
34	23	23
35	24	24
36	25	25
37	26	26
38	27	27
39	28	28
40	29	29
41	30	30
42	31	31
43	32	32
44	33	33
45	34	34
46	35	35
47	36	36
48	37	37
49	38	38
50	39	39
51	40	40
52	41	41
53	42	42
54	43	43
55	44	44
56	45	45
57	46	46
58	47	47
59	48	48
60	49	49
61	50	50
62	51	51
63	52	52
64	53	53
65	54	54
66	55	55
67	56	56
68	57	57
69	58	58
70	59	59
71	60	60
72	61	61
73	62	62
74	63	63
75	64	64
76	65	65
77	66	66
78	67	67
79	68	68
80	69	69
81	70	70
82	71	71
83	72	72
84	73	73
85	74	74
86	75	75
87	76	76
88	77	77
89	78	78
90	79	79
91	80	80
92	81	81
93	82	82
94	83	83
95	84	84
96	85	85
97	86	86
98	87	87
99	88	88
100	89	89
101	90	90
102	91	91
103	92	92
104	93	93
105	94	94
106	95	95
107	96	96
108	97	97
109	98	98
110	99	99
111	100	100
112	101	101
113	102	102
114	103	103
115	104	104
116	105	105
117	106	106
118	107	107
119	108	108
120	109	109
121	110	110
122	111	111
123	112	112
124	113	113
125	114	114
126	115	115
127	116	116
128	117	117
129	118	118
130	119	119
131	120	120
132	121	121
133	122	122
134	123	123
135	124	124
136	125	125
137	126	126
138	127	127
139	128	128

SEQUENCE DESCRIPTION SEE NO. 29b

[Faint, illegible markings]

CCGCGATAGG ACAGGGGCGAG CAAGCCACCG GCAACGCGGG CCACATCGCT GGACACCTCG	420
AGACCGTACT GCACCAACCT GAAGAGCTGA AACTCGCGG AACGTGCAAC AGCTGCGAAC	480
AATTGGG	487

(2) INFORMATION FOR SEQ ID NO:297:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 528 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:297:

ACGAAGCGCG AGAATATGAG CCGGGGCAAC CCGGCATGTA CGAGCTTGAG TTCCCGGCGC	60
CTCAGCTGTC GTGCTCGAC GGGCGTGGTC TGGTGTGGGT GCACGCTTTC GAAGGTTTCT	120
CCGACGCGCG CCATGCGATC CCGCTGGCGG CCGCCACCT CAAGGCGGCG CTGGACACAG	180
AGCTGGTCGC GTGCTTGGCG ATCGATGAAC TACTGGACTA CCGCTCGCGG CGGCCATTAA	240
TGACTTTCAA GACCGATCAT TTCACCACT CCGATGATCC TGAGCTAAGC CTGTATGCGC	300
TGCGCGACAG CATCGGCACC CCATTCTGTC TGCTGGCGGG TTTGGAGCGG GACCTGAAGT	360
GGGAGCGGTT CATCACCGCC GTCCGATTGC TGGCCGAGCG CCTGGGTGTA CGGCAGAACC	420
ATCGGCGTGG GCACCGTCCC GATGCGCGTT CCGCACACAC GACCGATCAC GATGACCGCT	480
CATTCCAAAC ACCGCGAGCT ATCTCGATT TTCAACCGTT CGATCTCC	528

(2) INFORMATION FOR SEQ ID NO:298:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 610 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:298:

TTAAGCGGCT TAAAGAGCGG TTGCGCGCTT TGGCTGCTGT TCGGCTGAGT TCGGCTGAGT	60
TTGCTGCTGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT	120
TTGCTGCTGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT	180
TTGCTGCTGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT	240
TTGCTGCTGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT	300
TTGCTGCTGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT	360
TTGCTGCTGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT	420
TTGCTGCTGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT	480
TTGCTGCTGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT	540
TTGCTGCTGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT	600

244

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:299:

```

Phe Asp Gly Tyr Glu Tyr Leu Phe Trp Val Gly Cys Ala Gly Ala Tyr
 1           5           10           15
Asp Asp Lys Ala Lys Lys Thr Thr Lys Ala Val Ala Glu Leu Phe Ala
 20           25           30
Val Ala Gly Val Lys Tyr Leu Val Leu Gly Ala Gly Glu Thr Cys Asn
 35           40           45
Gly Asp Ser Ala Arg Arg Ser Gly Asn Glu Phe Leu Phe Gln Gln Leu
 50           55           60
Ala Gln Gln Ala Val Glu Thr Leu Asp Gly Leu Phe Glu Gly Val Glu
 65           70           75           80
Thr Val Asp Arg Lys Ile Val Val Thr Cys Pro His Cys Phe Asn Thr
 85           90           95
Ile Gly Lys Glu Tyr Arg Gln Leu Gly Ala Asn Tyr Thr Val Leu His
100           105           110
His Thr Gln Leu Leu Asn Arg Leu Val Arg Asp Lys Arg Leu Val Pro
115           120           125
Val Thr Pro Val Ser Gln Asp Ile Thr Tyr His Asp Pro Cys Tyr Leu
130           135           140
Gly Arg His Asn Lys Val Tyr Glu Ala Pro Arg Glu Leu Ile Gly Ala
145           150           155           160
Ala Gly Ala Thr

```

(ii) INFORMATION FOR SEQ ID NO:300:

1. SEQUENCE CHARACTERISTICS
 A. LENGTH: 161 amino acids
 B. TYPE: amino acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

11. MOLECULE TYPE: protein

xi. SEQUENCE DESCRIPTION: SEQ ID NO:300

```

Arg Arg Arg Asp Leu Ala Gly Glu Leu Arg His Cys Ile His Thr Pro
 1           5           10           15
Thr Ile Ile Asp Gln Ala Asp Ala His Asp His Arg Thr Gly His Gln
 20           25           30
His Arg Gly His Ala Gly Gly Ile Asp Glu Pro Pro Gly Glu Cys Arg
 35           40           45
Glu Leu Gly Gly Lys Leu Asp His Ala Arg Asn Ala Gln Glu His Arg
 50           55           60
Glu Thr Thr His Pro Thr
 65

```

245

```

      100      105      110
Leu Thr Gln Arg Asp Thr Gly Ala Ala Ile Gly Gln Gly Gln Gln Ala
      115      120      125
Thr Gly Asn Ala Gly His Ile Ala Gly His Leu Glu Thr Val Leu His
      130      135      140
Gln Pro Glu Glu Leu Asn Thr Arg Arg Thr Cys Asn Ser Cys Glu Gln
      145      150      155      160
Leu

```

(2) INFORMATION FOR SEQ ID NO:301:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 175 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:301:

```

Glu Ala Arg Glu Tyr Glu Pro Gly Gln Pro Gly Met Tyr Glu Leu Glu
  1           5           10           15
Phe Pro Ala Pro Gln Leu Ser Ser Ser Asp Gly Arg Gly Pro Val Leu
      20           25           30
Val His Ala Leu Glu Gly Phe Ser Ser Asp Ala Gly His Ala Ile Arg Leu
      35           40           45
Ala Ala Ala His Leu Lys Ala Ala Leu Asp Thr Glu Leu Val Ala Ser
      50           55           60
Phe Ala Ile Asp Glu Leu Leu Asp Tyr Arg Ser Arg Arg Pro Leu Met
      65           70           75           80
Thr Phe Lys Thr Asp His Phe Thr His Ser Asp Asp Pro Glu Leu Ser
      85           90           95
Leu Tyr Ala Leu Arg Asp Ser Ile Gly Thr Pro Phe Leu Leu Leu Ala
      100          105          110
Gly Leu Glu Pro Asp Leu Lys Tyr Glu Arg Phe Ile Thr Ala Val Arg
      115          120          125
Leu Leu Ala Glu Arg Leu His Val Arg His Asn His Arg Pro Gly His
      130          135          140
Arg Pro Asp Glu Arg Ser Ala His Thr Thr Arg His Asn Asp Arg Ser
      145          150          155          160
Phe His His Pro Gly Ala Ile Ser Arg Phe His Ser Phe Asp Leu
      165          170          175

```

(2) INFORMATION FOR SEQ ID NO:302:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 175 amino acids
 (B) TYPE: amino acid

[illegible]

A: LENGTH: 921 base pairs
B: TYPE: nucleic acid
C: STRANDEDNESS: single
D: TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:302

AAATTGGGAAT	GATGAGCACT	AACAAGGGGT	TATTCATGCT	AGGCAAGCTC	ATTACGGGTA	12
TTTGGCAACAT	TGGCAGCTTC	AACAAGGGCA	TATTCATCT	AGGCAATTC	AACACCGGGG	13
AAATTGAACTT	AGGCAAGCTAT	TACAAGGGGA	TTTGGAAACA	AGGGCGATT	TACACCGGGC	14
GGTTGATCTT	TGGCAGCTAC	AGCAAGGGGT	TTTGTGAGT	GGAAATTATC	AGGGCTCATT	15
GGTTGGACCT	GGCTTTGGA	ATGCTTGAG	TTAATTGAA	TGCTTACAA	GGTTGGCGGC	16
GGATGGAAG	TGGGTGTAT	GATTAAGTTT	AAGCGCTNAN	CAATAACTAG	CATAACCCCT	17
TGGGGCTCT	AAATGGGTCT	TGAAGGGTTT	TTTGTGAAA	GGAGAACTA	TATCCGGGTA	18
ATTGGCGGTAN	TACGAAAAGC	GGCAGCGATC	GGTTGGCAA	GAGTGGCGCA	CGKGAATTGGC	19
AATGGACCTC	GGTTTACCG	GGCATTAACT	GGGGGGTGT	GGKTTTACCC	CCACGTNACC	20
TATAGCTCTC	TANNAGGCTN	RSGGCTATT	GGTTTCTTT	TTTCTTTCTT	CGCTTTCGGC	21
GGTTGGCTT	AGGCTAAAT	GGGNNNNNN	TTTMMGGGTT	TAATAATTCT	TTACNGSCCC	22

(2) INFORMATION FOR SEQ ID NO:304:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1082 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:304:

```

AATTCGGCAC GAGATANGGG CGCACCAGGG TCCGCAGCCG GCGGGACCGT CGCCAGCACC      60
ACCGGGGTCA ACAGCACCAC GGTGGCGTCC ANGCGAGCGG CCGCGGTGAT GGCGGGCTGAG      120
ACGGCRAACA CCTGCCGTAG CAGTCGGTGC GACTCCCCCG TCGCTGAGNC CATGGCCCGG      180
CGGGGTGGCT CGAACANGCC TTGGTCGTCC ACAGCTTAGC CAGGCANCCAA ACCGCAACCA      240
GAAACCCACA CGCCCGCGCC CCGGGANACC TCCGCCATCG KCTGCTGGGG CGANATCCCC      300
CGATCGCTNA CANGATGACC GGTGGCGGAA CCGCGCCGCT GCCTCGGGGG AGCGGCTTGG      360
GCGGGGCAAC CGCGAACCCA NGAACACGGC AAGCAGTATC ANCGCAACAG CAATTGTCAA      420
GGGCTAAACG CTTACATACC AGGGATCTCG CCGCGCCACA CCGTCGGMTC TGCAGSGCGA      480
CCCCNTCCTN GGGCGGNCAC TCNTCAAAGA TGCGATCNA CAGKCTAGGT CTTCGGCCGA      540
TATGSAAGGN CCCAACCGNT TTAAAGCGGC SAAAAAATC TCCCANTGGA TAAAAATCAGC      600
CGGGGANCCC CCGGTGSCMM NGTCYCGGKC ATTNTTCAAC MGGTTTNACG GCGGKTGCGG      660
GCGAACTKGC CAAAMTTAAG KTNNGGGGNTY CGGGCGCGTA ACCGGCINTK NGCCCCCTTAA      720
AAAACCGGNC YTTCTCKGAT TAMMACCGGN CCGCCAWTGG CGGKTCKTCC CANGNTYAAC      780
AMCCYCCGSS MNGSGKTGGG SAACCTTCC TSNNGGGGTTT NTKGTTSCYT AWMCCCCCGG      840
AAACCSGKYG GGKTGGCRTN WASSAMNCGG CMNGYTTCTT TAAAGGCCAN KNRAAWGKYT      900
CGTGGGGAAW CCTTCAATYC GAAATCTCTC CTTCMGSSCN CTTKQWRTYN NRNGGGAACS      960
AMWTTTCGTC GWTTCAWTCC GGTCCGASMA AAACKCTTTH TTTTCGSSSC STCCMGSSNC      1020
GGTGNANAN AAATATTTTC TCCNNANKK TTYCGGCTT TTKMGRNRNR GNGAACCGG      1080
35

```

(2) INFORMATION FOR SEQ ID NO:305:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 491 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:305:

```

AATTCGCAGG AGTATATGGG GTGAAGCTTT TAGGCGCGGT GGTGAGGTT GTTTGCGAAG      60
GAAATCGGCT CGANGTGGTT TCGTAGGCGG GTTCCANAA CGGTGGGGG GTGCGGCGG      120
ATCTGATCGG CGCGGCGGTA GTGACCTTC CCGGGCTTGT GCACTCGCAT TCGGAATG      180
TCTGTCTGCT GGTGTACCA GTGAAGAAAT GTTCTGACT GACCGGGTT TCTGTGAT      240
GATCGAAGC GTTTGGGAA GTGGGCTT TAATGAAAT TCTGAAAT TCTGAAAT

```

AWTCCCAACN	CCCKCCAANA	RCYKGGGGCC	CCNCCCAACC	CGGGKGAAGA	WTAATTTAAA	660
CCCYAACMAW	ACTWMMNACC	CNNGGGGCCY	AAMCGTYYNR	AGGTTTTTCT	NAAAGAAASA	720
ANTCGGAAMC	CGGNTSTACC	AAAAASCCCK	CCNWTCCCTC	CRASATTGSC	NCCSAAWKSA	780
AKGCCCCCN	TCSGCNWNNC	CSGCGGKKKT	KKGTTNCCCT	WMRCWMWYTS	GGCCNASCCN	840
CKYYSSMYCC	CCCCTCCCCM	CTCCGNKTCC	CCAMCCYANC	MGGCCCCCYM	GKCCCCWKNT	900
YKCCCCCCCC	AMMNNNGGGG	WGACCCCTGG	CCCCMKRRGM	TCCCNANTGA	MCCTCWGNRA	960
MKCYCCNRAR	ANMCCSCNCC	NGCNCRCNN				990

(2) INFORMATION FOR SEQ ID NO:306:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 223 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:306:

AATTGGGGTG	GCAACGGCGG	CCTGTTGGGC	AACGGCGGGG	CCGGTGGTGG	CGGTGGGGCT	60
GTTGGTGGCG	CCGGCGGGCG	GGGCGGTAAC	GCGGGGTGGT	TTGGTATGGG	GGGCGCTGGC	120
GGGTGGGGTG	GTGTANGTGC	GGCGGGGGCG	AACGGTGCTA	CGCCCGGTCA	GGATGGGGCG	180
GCTGGTGTGG	CCGGGTGGGA	CRACRGTCTG	GCGGCTGGTG	CCG		223

(2) INFORMATION FOR SEQ ID NO:307:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 418 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:307:

AATTGGGGTG	GCAACGGCGG	CCTGTTGGGC	AACGGCGGGG	CCGGTGGTGG	CGGTGGGGCT	60
GTTGGTGGCG	CCGGCGGGCG	GGGCGGTAAC	GCGGGGTGGT	TTGGTATGGG	GGGCGCTGGC	120
GGGTGGGGTG	GTGTANGTGC	GGCGGGGGCG	AACGGTGCTA	CGCCCGGTCA	GGATGGGGCG	180
GCTGGTGTGG	CCGGGTGGGA	CRACRGTCTG	GCGGCTGGTG	CCG		223

(2) INFORMATION FOR SEQ ID NO:308:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 418 base pairs
- (B) TYPE: nucleic acid

(2) INFORMATION FOR SEQ ID NO:309:

(A) LENGTH: 1036 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

X1 SEQUENCE DESCRIPTION: SEQ ID NO:309:

AATTGGGCAC	GAAGGAATCG	AGAATCGG	ATGCGTGAAG	CTCGGTCG	TGCGGTTACG	90
GAAGAACTCA	GGGTAGCGG	CGCGCGGTG	GAATGCTGA	GGCAACCGG	GAAAAGGGTG	120
AGGGCTGGGG	TGGAATAACT	GAANGTACT	GGATGGA	ACCGGTTAT	GATATGTATT	180
GGGCTGATCA	ANGTTTSTGG	AATGGGGAA	GGTGAAGGG	GACCTGTTGG	ATTTGGGGAA	240
TTGTATGGA	CAKACWGGC	CAGCGCGGT	GATGGTTGG	TTGAATTTT	GTGGCGGCCA	300
TANGTGGATG	GAATGATTT	TGATGGG	GATGAATA	TTGGGTATCG	GTACCGGAA	360
TGAGATTCG	GGACCTTCA	TGGCGGGA	GAAGMASGT	CGGANGTAH	GGTTGGCTTT	420
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	480
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	540
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	600
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	660
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	720
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	780
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	840
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	900
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	960
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	1020
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	1080
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	1140
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	1200
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	1260
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	1320
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	1380
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	1440
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	1500
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	1560
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	1620
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	1680
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	1740
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	1800
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	1860
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	1920
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	1980
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	2040
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	2100
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	2160
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	2220
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	2280
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	2340
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	2400
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	2460
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	2520
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	2580
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	2640
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	2700
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	2760
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	2820
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	2880
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	2940
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	3000
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	3060
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	3120
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	3180
ATTATGATCG	GGATTGGGA	AGTTGATCGA	GGGCTGAY	MTGATSGCGC	NACNCGWCG	3240
ATTATGATCG	GGATTGG					

(A) LENGTH: 1036 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:310:

AATTCGGCAC	GAGATCATGA	ATAGCGGGCT	GGTCAGCACC	GAAGTGGTCG	GCGATCTCGC	60
GAGCAAGTCT	CGTCTGCTCG	CCCAGCAGGA	GGTCGGGCATC	GATGCGGACA	CCTGCCGATGT	120
CTTGGATGGT	GTTCAAGTTC	AGGTAAGGCC	GACGCCGCAG	CTTTGCTAGC	AGGGTGTCTT	180
GGCTCTTCGC	ACGTGAGGTA	ACCAATAACT	CCGACGCAGA	CCAACTCCGG	CCCTCGATCC	240
GGGTACCAGG	CTCCGCCCGA	GCCAGCCGTT	GTGCCCCCTG	GGCCGAAGGT	CAGCTGCTGT	300
GCGATCGAAG	TAAGAAACCG	CCCCATGCCC	GTCCGCCAAGT	ACGACTGACC	GAGCAAACGA	360
ACGATCGTCC	TCCTTTCCGT	GGGGGTAATC	GAGCCAGCA	ACGCGACGAC	CCACCAATCA	420
TTGGGATTTC	GCCACTGACC	GACCAACCCG	CTGTGCGACA	CCCCAGCCGA	ATTGGTGGTC	480
TTCCGCGGGG	CCGNAACCG	AATCANCCGG	ACGCGCTCCG	CGAASCANCC	GCATANCNT	540
ACATANCAAC	CGNNTCTCCG	CCACATTTTC	GGGGTMTTGC	CCCTCNGCAA	CSSNAAYNCC	600
CCCAATTCVG	AACNAAAAAA	TTGGYCCATY	ARNGTYCTCM	CCAAAAACCN	AWTCCCEKTA	660
TCCCCCGGGG	GCGGCCCTYY	NMNAAAACCG	CCWAAANCC	CCSGGSCSCC	CGGGTTRWTN	720
CCCCCTGTCC	GCCCNCCSGG	TTTGGTCMCM	GGSCMTNWN	GGGNTGSCCC	CCCNCAAAAA	780
AAAAAYCKIG	NCAAATYAAA	CCCKYMAAAA	ASKTGGGSSC	CCCMARCCGG	GGKAARKWWA	840
ANTTAANCCN	KAAAAAAWW	NCANNMCCCC	NGGCNCCTAA	GGKYTTAGGG	GTTSTTNANG	900
ARAAAATMTC	CANATMSSK	TTNAAAAAAA	ASCCSWAKCC	CCCNHNKKN	CCAAWKAARP	960
SRCTTCGGG	TWNHSGGGG	KKKKTNCMS	KNNMTTWGR	CCCNCCGCTN	NNTWKCCCTN	1020
TCCTTGGNCC	RNCAGN					1036

(12) INFORMATION FOR SEQ ID NO:311:

1. SEQUENCE CHARACTERISTICS:

A. LENGTH: 1060 base pairs
 B. TYPE: nucleic acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:311:

AATTTGGAAT	AGTGGATTC	GATGGAACAA	AGCGACAGCT	ATGAGATTA	GATGAGAGCG	60
AGATGATTA	AGGCTAGTC	GGCAACCCG	AATGATTTTC	AGTTCTGAG	GAAGAGAGAT	120
AAAAGGGAAT	AGGCGGATC	GATCAACCCG	AGCAACCCCA	ATGGAAGATA	AGGTTGAAAT	180
AATGAGACAA	TGGCGAAGG	AGCTTGACAG	GGCGCGCACC	ATGCAAGCTG	TTAGAGCTGT	240
AGTTCTTGA	AGAAGCGGGT	TGGCCACCCA	AGATCAGGCT	GGCAAGGGG	ATGAGATCAA	300
TGTTCTGCTG	ATATCGGCT	AGCTTGGGCG	GGCGAAGAA	ATGAGCTGAG	GCATTACGAT	360
AGCTTCTGCT	ATGACTTTTC	GACAGCTGCG	CACCAAGACT	ATGAGAGGCT	TTATGAGGAG	420
TGCTATGCA	ATGCGAGGCT	GCTTCAAAAA	CTGCTTGTCC	ACAATGCTAT	TGTTGAGTCC	480
CGAATTCCT	ATGCTTGA	GAAGATNCA	TGTTNCGGCT	NAACAACTG	GTGTAAGAA	540
AGAGCAATA	TTGAATGCT	ATGCTTGA	GAAGATNCA	TTGAAGCTG	TGGAAGAA	600


```

AAGGKKKTKC SCMACCCCAA AAANMMNNCN AWHCCCGMGK SARGGGGRNY TTMEAGGGMG 960
GNYCCCCCW YCGGGGGGNA NAAYAAAAGK NGSNGRGAAT NTNTNTTTTGK RSSSRNKTTT 1020
TYNTCTTCN CCMGNRWWG SRAMNTGKTS NSSGGGSGGC 1060

```

(2) INFORMATION FOR SEQ ID NO:312:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1040 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:312:

```

AATTCGGCAC GAGCTTCACC AAAGAGCTGA CATGCGGGGT BATGCGACAT CGCATCGAGG 50
GCATACCGG CATGGATGAN CCGAANGGAN TGTGGGGTTC GGTCAACTGG ATTACGGTTC 120
CCAGGCTGAA ACGCTTTGGG SCGAAAGATG CGACGCTTAA GTTGGGCTTC CACCGTCCAA 180
TGTTNGTATG GATGCTGGAA CCGCGCTGAC NGATAANGAA TTGGCTGGT GCGGGGCGACN 240
ATGGATGGTC CKSTTTTCNC TCCGCGGTTA AATTGGSTGT GCATCATCTG GCAGGCTATG 300
TTCCCGCTAC RCTGCAGGCG ATCATGGATG TGCGGCTAAC GAANAAGTTA TGACATGGCG 360
CAAGCGAMTC GGGCATSCNC GCGGCAMTTT CGCAACCTGC TGTGTNTGAA GCGTMTCAAC 420
CGAATCCGGC GCTYAAAAGC NGGCTTGGCT TGATTMMAAC CNAACCCNTN CNATYCTTTG 480
CCGNGMNTG CGTTCTCTCC AACTCGGKKG SYTGCCNCCG TGAAACCCMA CTNCCCCCCC 540
GTTGGACTTA MRTNTTCAAA AAMCGGMTNA ACCCGAATN SAACCTNCCR TCAAANTAMM 600
SAANTCGGGC TTYGGGNRCC CCGGGAAYW TTGNCNGGG GMNTYCTCN GGTTYNGGG 660
SAAACNTTTC CORTNCKMN TTIACAMGGC NCMTNMTTGM GGGSCSNNAS GWCCCCGGKK 720
TNTTTCNAW TONENSKTTT TTKGGGGGGG GGGYGRTRMC NCGGGCCCCC GGCCCKKMAA 780
AAAAAMCMSA RRCNCGGGG KKCCGCGCCX NMATNGGGGC YKCRAAACAA ACCCCANRA 840
TNGNGMGGGC SMACCGNGN SYHAAAGGT TSNSTMANM MKGMANNCT SGMSCCMNSN 900
NCTGMSGGKT TTKGNGARN AANAMKMGSM RSGNCGGCRN GAAAGGGGMS GSCKSCNNGN 960
NGASNGWMN TRNGANRCC NONGYBMPN BRUNNGNBNN CGRKNNAON NMKMCANSMC 1020
NSRNGNNNN CGYMTNKGCC

```

(2) INFORMATION FOR SEQ ID NO:313:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 148 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:313:

```

AATTGCTTAT GAGAAATAT GGTCAACTGG BATGCGACAT AATCGGGGTC GGTGGTCACT 50
GAGCTTTCAT TCAACTTATC TGTGGGGTTC GGTGTATG GYLLTATAT ATGTGGGCTC 120
ACGCTTAAAT GATGCTGAA TTTGCTGGAC NGATAANGAA TTGGCTGGT GCGGGGCGACN

```

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 332 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:314:

AATTCGGGCAC	GAGAGACCGG	GTGTTGACC	AACGGACGCT	TGGGCGCGGG	CCCTTGGCGT	60
GGCATCAGCC	CTTCTCCTTC	TTAGCGCCGT	AACGGCTGCG	TGCCTGTTTG	CGGTTCTTGA	120
CACCTTGCGT	ATCCAGCGAA	CGCGGATGA	TCTTGTAGCG	CACACCAGGC	AGGTCTTTCA	180
CCCGGCCGCC	GCGCACCAGC	ACCATCGAGT	GCTCCTGCAG	GTTGTGGCCC	TGGCCGGGAA	240
TGTACGCCGT	GACCTCGAAC	TGACTCGTCA	CTTCACGGGG	GCAACCTTCC	GAAGCGCTCA	300
CTTCGGCTTC	TTGGGAGTGG	TGGCTCGTGC	CG			332

(2) INFORMATION FOR SEQ ID NO:315:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 962 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:315:

AATTCGGGCAC	RAGTCGTCT	AGACGGATTC	AATGCTCCCG	CGAGCACCTC	GGCACTGCAC	60
ACCTTCGAGC	AAAATGTGCT	CAATGTGGTG	AACGAGCCCT	TCCAGACGCT	CACCGGCCGC	120
CGCTGATCG	GCAACCGCCG	CAACGGGACT	CCTGGAACCG	GGCTGACCC	GGGGCCGGCG	180
CTGCGCTGTT	CGGCACCGGC	GCGAACGGCG	GCTCGGGGGC	GAACGGAACC	AACGGCGGGG	240
AGCTGGGJAC	CGCCCCGGCG	GGATTTCTTC	GCAAGGGGGC	ACCGCGGGGG	CCGGCGGTGT	300
CGCAAAAGCG	CACCGCGGGG	GACCGNGCCG	CGTNGGGCG	GCTTCTKGAT	GGGCTCCCGC	360
GGTNACGGCG	CACGGCGGGG	CGCGGCTCAC	CGCTNGTTGG	GACCGGGGGA	CGGCTNACCC	420
CGATCTTCTT	CGCNCCTCG	GAAACTCGCG	GGCGGCCCC	ACATTAKACC	CGCGCGNACC	480
CGGMCCTGG	CGGAACGGNG	GGTNTTTTCC	AACGGGGGGG	CGCGGAACC	GNMGGSTGTT	540
CGTNGGGGGA	AGGNCBAAT	CGGKCTANG	TAATTCCTCG	ANGSKTGAMC	CTBATSSNCA	600
CTTGMAGGAA	CTTNCBANT	CTTSPACAT	CTTNGGAAA	ASPAWNKNGT	CGGCAAACTA	660
CTTCTTTTCT	NATKGSNNA	AAAAGCTCT	CTTNSPACT	CTTNGGNGM	CTTCTTNTTN	720
CTTCTTCTCT	CTTGGGNAAM	CTTCTATTTG	CTTNGGNTTN	CTTCTMNNNA	AACTCTAAM	780
CTTCTCTCTCA	ANGGGKSNCG	CTTCTMNSGT	CTTCTCTMRA	CTTCTCTCTCT	CTTCTCTCTCT	840
CTTCTCTCTCT	CTTCTCTCTCT	CTTCTCTCTCT	CTTCTCTCTCT	CTTCTCTCTCT	CTTCTCTCTCT	900
CTTCTCTCTCT	CTTCTCTCTCT	CTTCTCTCTCT	CTTCTCTCTCT	CTTCTCTCTCT	CTTCTCTCTCT	960
CTTCTCTCTCT	CTTCTCTCTCT	CTTCTCTCTCT	CTTCTCTCTCT	CTTCTCTCTCT	CTTCTCTCTCT	962

(2) INFORMATION FOR SEQ ID NO:316:

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:316:

AATTCGGCAC	RAGAAGACCC	CCGAANGTTT	GCGCTGGCTC	TACAACTTCA	TCAARGCGCA	60
GGGGGAACGC	AACTTCGGCA	AGATCTACGT	TCGCTTCCCC	GAAGCGGTCT	CGATGCGCCA	120
GTACCTCGGC	GCACCGCACG	GCGAGCTGAC	CCAGGATCCG	GCCGCGAAAC	GGCTTGCGTT	180
GCAGAAGATG	TCCITCGAGG	TGGCCTGGAG	GATTTTGCAN	GCGACGCCNG	TGACCGCGAC	240
GGGTTTGTG	TCCGCACTGC	TGCTCACCAC	CCGCGGCACC	GCSTTGACCT	CGACCAGCTG	300
CACCACTCGT	GCCGCTCGTG	CCG				323

(2) INFORMATION FOR SEQ ID NO:317:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1034 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:317:

AATTCGCAGT	GTGTGTGGCG	GCCTCCAGAA	GAAGATGATC	GCGAACATCG	CCAGCGCCCG	60
CCAGGCTATG	GTGCCGGTGA	TGGCCGACCA	GCAGATCATC	AGCGGCATAC	AGCCGGCCCG	120
CCCACCCAC	ACCACSTTCT	GTACCTGCG	TCCCTTGAGC	CAAAGCSTGT	AGACRAACAC	180
ATAAAACCG	ACGGTACCA	GGCCAGCAC	CCCGCCAGC	AGGTTCGTGG	CGCACCATAG	240
CCAGAAGAAC	CAGATCACCG	TCCACCTCAC	CCAGTGCCTA	ACCGCTTTCC	GGTCGGCACC	300
CCCTCCCGCG	CCAAGCGCGG	CCCGCGCGTT	CGCTTCATCA	CGTTGTGAT	ATCGGCGTCC	360
GCACCACTTT	GAGCGTCTTG	CCCGCGCGCG	CCCGCATCAT	CCCGCGCGACN	ANCGTGTGTA	420
GCATGANCAG	CGGATCAATG	CCCGCGCGCG	TCTGCGCGCT	CGTCCCGAAT	TCAACTCCCT	480
GCACCACTTG	CCGCGCGCT	CGAAGCCCGG	TGAATGAWTG	AATTTAAACC	GSTSAACANT	540
AACATACATA	CCCTTCCCGG	CTCTTAACCG	CTCTTGAANG	GGTTTCTTCC	TTAAAGGAAG	600
AACHATTTTC	GCATATCTCG	CTCTTWTATC	GAAGAGGCCC	CCCGCATNGC	CGTCCACAGT	660
CTSCCCCTGA	ATGCGAATCG	CCGCGCGCT	CCGCGCGCT	CCGCGCGCT	CCGCGCGCT	720
CCGCGCTGA	ATGCGAATCG	CCGCGCGCT	CCGCGCGCT	CCGCGCGCT	CCGCGCGCT	780
CCGCGCTGA	ATGCGAATCG	CCGCGCGCT	CCGCGCGCT	CCGCGCGCT	CCGCGCGCT	840
CCGCGCTGA	ATGCGAATCG	CCGCGCGCT	CCGCGCGCT	CCGCGCGCT	CCGCGCGCT	900
CCGCGCTGA	ATGCGAATCG	CCGCGCGCT	CCGCGCGCT	CCGCGCGCT	CCGCGCGCT	960
CCGCGCTGA	ATGCGAATCG	CCGCGCGCT	CCGCGCGCT	CCGCGCGCT	CCGCGCGCT	1020

(2) INFORMATION FOR SEQ ID NO:318:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 101 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

```

AATTCGGCAC GAGCCACAT CCGGGGCGCG TCSTTGCATG ACTCGTTCGT CATCGTCGAC      60
RAGGCACAGT CGCTGGAGCG CAATGTGTTG CTGACCGTGC TGTCCCGGTT GGGGACCGGT      120
TCCCGGGTGG TGTGACCCA CGACATCGCC CAGCGCGACA ACCTGCGGGT CCGCCGCCAC      180
GACGGGTCCG CCGGGTGATC GAGAAGCTCA AAGGTCATCC GTTGTTCGCC CACATCACCT      240
TGCTGCGCAG TGAGCGCTCG CCGATCGCCG CGGTGGTCAC GAGATGCTCG ANGAGATCAC      300
CGGGCCGCGC TGAGTGGGCC TCCCGCGAGC A

```

(2) INFORMATION FOR SEQ ID NO:319:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1026 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:319:

```

AATTCGGCAC GAGATGCTCA CCGTGGCGAC CAGTGCACCG AGGCACCGCC ACCAGTTACG      60
GCTGATGGGC CAGAAGATGG ACCAGGTGCT GCCCATCCCG CCCACCGCAC TGCAGCTGAG      120
CAGCGGGATC GCGGTGCTCA GCTACGGCGA TRA3CTGGTG TTCGGCATCA CCGCTGACTA      180
TGACGCCCGC TCCGAAATGC AGCAGCTGCT CAACGGTATC GAACTGGGTG TGGCGCGTCT      240
GGTGGCGCTC ANCGACAATT CCGTGTGCTG GTTTACAAGG ATCGGCSTAA GCGTTTCATCC      300
CGCGCACTCC CCANCGCTGC GCGGCGGGGG CCGCCCTCTG TGGCGACCGC CCGAGCGCGT      360
CACTGACGCG ATCTCGCTCG GCSTTAACCC CGTGAGAAGG TGGGTGCTGC GCAAGTTGGG      420
CCCGGTCAAC ATCNATCCCG GCCGCGCATGA CCGNGT3CTG TTCCACACCA CNTSNGACNC      480
CCCCCAGGAA CTGGTCCGGC AMTNCAGGAA NTYCGT3TGG GCACENGCTT CTTCCGKTRT      540
GGCYTAAACT TCCNATSTTN CGGCGGGCCT CTGGCGTTNC GNCCGGGCGG NTCTTNCCAA      600
ATCGGSMMAA ATCCCCANMC AAACCGCCCG GGTCTT3GGG GCGGGGNGGC GGCNAWNCC      660
AAACCGCCCG NTAAANTCT TTGKTCNNN CNCGGGNCC NCNAANSCAN CCGTTTGGG      720
NCTTCCCGCC CCANWTTTAA CCGAKCGSCN AAYCGCAASY TMMGKCCYCY KNAAAAAAAA      780
AATTTGCGCG CCGCAANTAA ATTCCGNGGC CCGTTGGGGG CCRANCFYNT TTTMCCSNSS      840
TGNRRAAME NGGANCSSG KAAYTMTTNS NAAYCGCCSN AAMNTTTTC TAANNCCCN      900
TNCSSGAAA ATTNVAMAAM CMKKTGNSG GGGKTTGNC SGKKGRAGGM AAAAAANRSN      960
SKTTTMCNNN SANMNCNSNN GGGNSNNNNN NNNKNCYKNC CSNAANMCCC CCGCGGGGGG      1020
CCMMCC

```

(2) INFORMATION FOR SEQ ID NO:320:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 324 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:320:

GCACCACTCG TCGCGCTCGT GCGG

324

(2) INFORMATION FOR SEQ ID NO:321:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1010 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:321:

AATTCGGCAC	GANGCGTGCC	GCTNAACACC	AGCCCCGGGC	TGCCAGATAT	CCCGGACTCG	60
GTAGTGCCTC	CGGTGGGCTC	GTTGCTCTCC	TSACGGGGCG	CGGCGACCAT	AAGGTGCGTM	120
ATGCCCCAGG	AGCGGCGGAG	GTGCATGGAG	TGCATGATGA	TGCGACTCTC	CAGCTCGCCG	180
ACCGGGAGCT	TGGCATCGGG	CTGATCAGC	CAGGACGCGT	AGGACAAGTC	GATCGAATGC	240
ATAGTGGGCT	CCAGAGTGGC	CGTGCAATTC	CTGCGTGCTC	CACGGCAAAT	GCCTTGATTT	300
CTACTCGGCG	TANTGTTCCG	GCATCGGCTG	CGGGATGAAT	GGGAACCGCA	SGATGGCGAC	360
GAACGGGTCT	GANCTCAGGT	TTGCGGCTTT	GCGCACAGTG	GTCNACANCC	GCTACTCGGC	420
ATANATCTGG	CCCNAAATCG	GCGCGGACGG	CGGCCACNAT	AANAACGGGC	ACNACAATCG	480
CCGCCCCCGT	CACCCNAACA	ACANCTTGSC	ATCGGATTTT	GTCCCCANCG	CTCAANCCGT	540
CCCGAACGCC	TGNTCCGGCG	NACTTTTCTT	NNAWTAAGTG	CCGCTTCCGK	CCCTGGNGCA	600
WTAAATGGGA	AACCCCTNCC	CCACCTTGAA	GGGCTTGTTG	NATTTTFACT	GSTAACCCCG	660
AATTTTTCGG	GANTCGGTCT	KCCGGGTTTT	YSTNTTCCCC	ACCTTNGNAN	GGGCCGGCCA	720
AGSTTTTCTT	SVTGAAGGGG	GAAACCCAAC	TTTNTYTYYN	AACGSCMNAA	MYMTTTYCSG	780
MNAASCTNKT	CCCCCTTAAC	CAMGGSGGTN	AACCGKTMNG	NGGKTAAAAA	GGGSKNNKTG	840
NCCCTTMMANG	GGGGGAAAAA	TSTKTCNNCG	GGGCCAAAAW	ACCMMMMYGN	GTGKKKNNKSS	900
TCSAAATTTT	NMMRAACTTN	GGGGCCSSGA	NITTTNAAAG	MSCCCCCSHN	GSTGKCCCN	960
NTTTCNNMAA	NMKKGKQWNN	SNMNSCGNGG	SKQNSGSSNN	NNAAGMGGGG		1010

(2) INFORMATION FOR SEQ ID NO:322:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1010 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:322:

AATTCGGCAC	GANGCGTGCC	GCTNAACACC	AGCCCCGGGC	TGCCAGATAT	CCCGGACTCG	60
GTAGTGCCTC	CGGTGGGCTC	GTTGCTCTCC	TSACGGGGCG	CGGCGACCAT	AAGGTGCGTM	120
ATGCCCCAGG	AGCGGCGGAG	GTGCATGGAG	TGCATGATGA	TGCGACTCTC	CAGCTCGCCG	180
ACCGGGAGCT	TGGCATCGGG	CTGATCAGC	CAGGACGCGT	AGGACAAGTC	GATCGAATGC	240
ATAGTGGGCT	CCAGAGTGGC	CGTGCAATTC	CTGCGTGCTC	CACGGCAAAT	GCCTTGATTT	300
CTACTCGGCG	TANTGTTCCG	GCATCGGCTG	CGGGATGAAT	GGGAACCGCA	SGATGGCGAC	360
GAACGGGTCT	GANCTCAGGT	TTGCGGCTTT	GCGCACAGTG	GTCNACANCC	GCTACTCGGC	420
ATANATCTGG	CCCNAAATCG	GCGCGGACGG	CGGCCACNAT	AANAACGGGC	ACNACAATCG	480
CCGCCCCCGT	CACCCNAACA	ACANCTTGSC	ATCGGATTTT	GTCCCCANCG	CTCAANCCGT	540
CCCGAACGCC	TGNTCCGGCG	NACTTTTCTT	NNAWTAAGTG	CCGCTTCCGK	CCCTGGNGCA	600
WTAAATGGGA	AACCCCTNCC	CCACCTTGAA	GGGCTTGTTG	NATTTTFACT	GSTAACCCCG	660
AATTTTTCGG	GANTCGGTCT	KCCGGGTTTT	YSTNTTCCCC	ACCTTNGNAN	GGGCCGGCCA	720
AGSTTTTCTT	SVTGAAGGGG	GAAACCCAAC	TTTNTYTYYN	AACGSCMNAA	MYMTTTYCSG	780
MNAASCTNKT	CCCCCTTAAC	CAMGGSGGTN	AACCGKTMNG	NGGKTAAAAA	GGGSKNNKTG	840
NCCCTTMMANG	GGGGGAAAAA	TSTKTCNNCG	GGGCCAAAAW	ACCMMMMYGN	GTGKKKNNKSS	900
TCSAAATTTT	NMMRAACTTN	GGGGCCSSGA	NITTTNAAAG	MSCCCCCSHN	GSTGKCCCN	960
NTTTCNNMAA	NMKKGKQWNN	SNMNSCGNGG	SKQNSGSSNN	NNAAGMGGGG		1010

AATTNTTCCG	GANTCGGTGN	KCCGGGSTTT	YSTNTTCCCC	ACCTTINGNAN	GGGCGGGCCA	720
AGSTTTTCTT	SYTGAAGGGG	GAAACCCAAAC	TTTNTYTYYN	AACCCSCMNA	MYMTTTYCSG	780
MNAASCCKNT	CCCCTTTAAC	CAMGGSGGTN	AACCGKTMNG	NGGKTAAAAA	GGGSKNNKTG	840
NCCCCYMANG	GGGGGRAAAA	TSTKTCNNCG	GGGCCKAAAW	ACCMMMYGN	GTGKKKNKSS	900
GCSAAATTTT	NMMRAACTKN	GGGGCCSSGA	MNTTTNAAAG	MSCCCCCSNN	GSTGKCCCN	960
NTTTCNNAA	WMKKGKNWNM	SNMNSCSNGG	GKYNSSGGSNN	NNAAGMGGGG		1010

(2) INFORMATION FOR SEQ ID NO:323:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1092 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

XI. SEQUENCE DESCRIPTION: SEQ ID NO:323:

NSNGGGGWNS	NTCAYCAYCA	YCACSGGGW	GNATTGCGGG	CGCAWCTTGT	MAASAGATCT	60
CSAAYTCGGC	AMGAGGGAMT	CKCTMGCNCT	GCTGTGCAAN	CCAATRAGGT	CTRATAATTY	120
CUACTCCACA	AAAAACCGTT	GTGTGTAYYT	SCCGRAAATR	AAGGCGCCGG	TNTCAACWYC	180
GCUGGTKTTY	CCRATYCCCG	TKTTGTAMCT	GCCKGGGTSR	AAAYCCCCGG	TGTTGGAYCC	240
CGGATTGAA	ACTSCCGGKT	TGAAACTGCC	GKTTTSGCSA	TCCGGKWATT	GAMSTCRCCG	300
ATTAAAAAAC	CGGKKTGGGN	GCTGSNCSTG	CCAAATNCGR	AYCCRATAYC	CCATGGCCCTG	360
KYCTYCTCK	YCGGTACCCA	AAVCTGGGTA	TGCTATACTG	GYCCCTAAAK	GCAAWYCKGG	420
GCTSYCMMTK	TTGEXKGGST	CCNAATTTAS	CACCASGGGT	TCCTTCGATA	CCNAAAACNCG	480
GKTGGGWCWC	AGMCCGRAAA	AAAKAATAAT	RAKAAKGGTG	CATNYCCAAA	ACCNCGCCCN	540
CCGNANTNEN	ATCCGNTNCC	MSCNCCCCCA	GGGGTNAAGK	TKSGGAAYTT	CTMMAACCCC	600
TAANCCCCCA	TAACNTNCGR	GAASAAAACC	CTYCNCGGGG	GYCNWNCAAA	ACASCTTTAT	660
TGGCTKSTTT	CGGGMWCCGT	GCCCCCNAAA	YCCCAAATA	CTTTTGTGGT	CCNAGAKAAA	720
ACGTCGGGGN	CCMGCCSNAA	NWTATYTCTT	KGGCAANCCG	CSAAACCTTR	TCMNACCCNK	780
ATPMTCCGTT	CCCTTSCAAT	TGGYCGGRAT	NCSSNCCTYT	TCAAARKKKS	CAKWNWNGNG	840
BRNNACOMA	ACCCCAAGTY	CCNAAAATN	JKCCCCGCTG	CNAACACGNK	TYTTCGAAA	900
ASCCCNCCCC	CCNCCCCRAA	AACCCCCCNA	CKANTNCCCA	AAAACNPNK	GGCCCCCCCC	960
CAAAACMAAAA	AMCCCCCGGM	AMACGGGGGN	NMCCCCGKKK	KKTTTTCTTT	TKCCMRSCCG	1020
AAMSCAMWSY	KSKTMAAAAA	GGAAURANCN	TYCCCANANM	TCCCNVWRSW	CCGSWGMGNA	1080
GAASMCCCCC	CS					1140

(2) INFORMATION FOR SEQ ID NO:324:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1280 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

XI. SEQUENCE DESCRIPTION: SEQ ID NO:324:

ATATTCCAGG	CACCACGCCC	AGTTTGGTGG	ACAATGCCCT	GGCAKTTTCC	TCRAANTTCG	300
TGAAACCGAA	TTTNSMTTGA	ACCNCCAARG	CCCCSNCCNR	AACARTTGGG	WTCCGCGGTT	360
CTCCCCACCG	KTTTCCGGGG	GTNTCGGCAN	AANCGCACCC	WTGGWTTCTM	TCNCCGCACC	420
GGGCGGACAA	NTCGGTTGC	AATTTGCRA	AYCGGGGCGG	GGATTCCSCA	AACGGGTGCC	480
GAAACTGTTY	YCRAAMACCG	GGAKCCGCAA	TTTCCGGGCR	ANAAATTTCTN	YCNCACCACT	540
GCTTCTACTT	CCCCGACCGT	AACMANTTTC	ATCGTCTNNN	CCTCTGCCCT	TGGGGCAGGG	600
CKAAAYACCG	CMTTKGGTTT	CGCAACCTGC	CGCCCAANTC	CCNAMCCRCA	CTTTCNATTT	660
GGNTCGAATT	SCCCCCCGGT	RANAACCSCC	NTGGCCNNYT	CGGASSAAAA	NGGGCCCTNT	720
KGGCNSCCCC	AGTAANACCC	TACCNAYTS	CAWTCTTTGC	CAAASITKGG	ACGAANSKTG	780
GGNTTCCGGK	ATTTYTTTGS	GGNCNCCCTN	TATNGGSNTN	GGGCCCKCYN	NCSTKTGKCA	840
NASSKAYCCS	NGNKGGGGGT	ACCCCCCTMG	GGGGGTTTTT	NSGGCCCCCC	AWAYGNKSTG	900
GCCCCCNNGG	GGAAKAATWT	MWWTMCNSGG	GGGAAWTTTT	NTSTGGAMCS	SGGACYCCCR	960
GGGGGKTTTT	TCCCCCNCSA	NNAWANGGGG	GGGGGANAYT	NTGNSGNGGG	KWNTTTATTT	1020
YTYCYCCTM	TKACMSGGGG	GTTTTKAKNG	GGGGGAGAAA	ANAAAAAAA	RAKGGYKNTT	1080
TSKNCACNCT	GKWNWNWANR	NAGAGTCCCT	CKCKCCNCSC	SNTTCTTTTT	MGNSGSYGGG	1140
GNNGNNMAAA	ACNKSRRMAC	KCSYTYCCCG	CGYCTCCTCC	NCNGGGGYGS	NGSGCNSTYN	1200
BNMKGRKMTA	TNTMGNCGTN	SCUTCCNCCC	GCKNKNTGTC	TMTCNMYGSG	C	1251

(2) INFORMATION FOR SEQ ID NO:325:

(1) SEQUENCE CHARACTERISTICS.

(A) LENGTH: 1099 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(12) SEQUENCE DESCRIPTION: SEQ ID NO:325:

AAATCGGCAC	MGAGTATCAC	CAAKCTGYGT	TGCCCCAGCAA	AGTGGAGCTA	TACTACCTG	60
TATGTGATCC	TCRACATCTT	CTCCCCGTAC	KTGGTCGGGT	GGATGGTGGC	CTCGCKTGAK	120
TCRAAGGTCT	TGSCCRAACG	GCTGATCGCG	CAAAACCTTG	CGCCCCAGCAC	ATCAKCGCCG	180
AACAGCTGAC	CTGCMCCCGG	ACCGGGGGYC	GNCAATAACT	CCAAAACCGGT	GGCMCTGCTG	240
TTGGCCNACY	CGGTSTGCCA	ANTCGAACTC	ASCCSGGNMA	CCAKMAACKA	NAACCGTTGT	300
TTGAAGCCCA	GTTCAAAAAC	CTCAAGTWCC	GGCCCRACCT	CCCGAAACGG	TNCGAGTCKA	360
TORSAGGSGG	CGCGGTGCMC	TGCAACCGGT	TCTTCGGNTG	GTRCAMCCCN	AAAMCAAGCA	420
TTCCGGGMTG	CGMMTCCCGA	CGCCGCCAAS	TTTMCCTACG	GCSGSCCNAT	CAAATTCGCC	480
TGGAACSGSN	CGHCCCTCNK	CGAMACGCCC	TWCCAAAACC	TYCGAACGGK	ATCCTTCKGY	540
NAACNCCCGA	RENCCCKSKT	TGCGGGCTTC	NMSGCGAATA	TCCCKNSCMNT	CGGAATCCAA	600
TTCCCMKYGA	TTTTTNTTCC	CGCCGGCCCG	AAAKNGGGYC	CTASSNMKC	ENCCAMNANT	660
GNWATCTGG	NGTTCCUNAN	TTTGGGCTTC	NMAATSAMNA	NMNRGGGTCT	TSCYACCMNT	720
AACCGKTHG	KCCCKKCTK	MANAAAKATT	RATCAMKWNG	GGNKCKCNEN	NAAMACCCCN	780
ENYNCWYTC	TYCCKSKWNT	TCGMYNANCA	NGCGGGAGGW	GGSGRMKMC	CTMTCTCNCT	840
MCGCGCTNTN	TYCKSGAKAT	ACASMKKTC	CGCCNGCCCN	MAAMANRAKA	CTAKCCGYGN	900
CCSNSTCTTN	CTSNMKKMR	TCGMMWNATC	NTTYGKKCN	KCTMKATNWC	TSCTSKCNCK	960
MRAMCTCTYG	SNMTCTCTCA	TGCTCKKCC	SNMSKCTCKC	KSCNCCNWN	CNKCNMKCN	1020
GGNSTCTCTY	TCTMKNCTG	AGCKCCSKNC	WACNCACACK	NGWCTTTTTT	WKNNNMKCNM	1080
TCCKCACCP	MTMTWCCG					1099

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:326:

GNGNTATACA	TCWCTGTGYA	CCSAGGATCW	ANTGCGGGCG	MAAKCTWSTM	CASAGATCTC	60
AAAYTCTGCA	MGAGCGGCAC	AKAKYSTCGT	CCMRACCCGG	CAYACWCCWG	CNCGCCCCWT	120
CTTRGACCGG	GGCKATASMC	ACCGTTGGCC	CCGGCNCGCA	CCTACACCAC	CCACGCCGCC	180
AGCGCCCCCW	TRAMCAAACC	ACCCCGCKTT	TACCGCCCGC	GCCGCCGGGG	CCACCACCAG	240
CCCCACCGGC	ACCACCGGCG	CCGCCGTTGC	CAAAACAGGC	CCGCKTTTGC	CACCRA	296

(2) INFORMATION FOR SEQ ID NO:327:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1073 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:327:

NONGSSNEMY	ATCATCWTTC	TGCACCSNGG	MTCWATGGCG	GCCGCAATCT	TSTMNASAGA	60
TCTCGAAYTC	GGCAGGARCA	TCTGCGCGGN	GAATGTCCAA	AWGTCKWTKA	CGGCMATCGG	120
TTTGCCGYCA	ACCACKCTRT	SCAKATGCGG	GCCAMWTYCA	AACCRATTAT	TTGGGYCGAG	180
AAAATTTMCG	CKTGTRASCA	ACCTGTAGCG	GSTCAASCAA	CAGCCTCTRA	ACCGTAAATY	240
CKTAGGTNKT	YCCGGCAACA	ASCYCRATAA	TSCGGCCCGC	AMCCACAAAA	CCTGANTNGT	300
TNTTCNCRAA	NCCGGTYCCC	GRAGGGGTSA	ACTGCGGTAR	GCTTNTCWYC	NCCTTRACAT	360
TAAACCCGCC	CGGNTCWTCS	CCGCGGCCAA	ATYCTTSCCC	WTKGCNACCA	YCCGANCCTG	420
CGGTATGGTS	RAANCASTSG	GCRAACGGTM	MRCSTACCHC	TGGGTGATYC	KTCGGNTCCG	480
BNAATTCGGG	GATTTACGGG	LAMGGTTAAY	CGAGGYCCGG	TNTGCTCKXY	CNACAACCCG	540
ATCMWCNCCG	TACCTKTTAA	AATTCCTTGT	SGTGGAACCG	AWYCKAAANA	NMTNTYCCCN	600
TCCAMMGGGG	CMCGGAAKYT	CNACNTGGKT	NACCCCTNCG	YTTGAASTTT	TCYTGNCCTT	660
GGGCCAAAS	AWACCSGAKT	CCCGGAAYCS	WTAGGCYTGN	TGCCCCSTTA	AATTGKNCYT	720
AATCCCKCAA	CGCTCCCCCG	GGTCCSCCMT	TAAAMTTCCG	CCCKSCASNG	GAATYCYKSG	780
CGNSTMATTW	CCNCCGNTTT	THYGHNAAAJ	SCCCCCWKN	TGCTHCCCN	GNTTGGGCTT	840
CTTTCGAMYC	AAAWNTNGGS	MMCNRAGNCS	SGNAMCCSTN	CKKSGGSAW	TKAAYTYGGR	900
GGGGSTCNYC	CCGRCGNAA	AAGYCTHGGJ	KCCGSSCCYT	CCWARTTTYT	CNGGMRCMAV	960
ATMANSGGNG	CTCCGSTNCW	NGCCTTCCCN	ANSMAMAAAN	WKKKCCGSGS	TCGARRNWNV	1020
WCTTUNGGNG	WTCCCKNKTG	NSCNSGNCSS	YGGNSASWTC	YNYCNCACCA	ANC	1080

(2) INFORMATION FOR SEQ ID NO:328:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1166 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

CGCCCCGGTTC	TTMMMTTCAY	TCATTCACCG	GGMTCTAGTG	CGGCCGCAAK	CTTGTCACAA	60
GATCTCGAAY	TCGGCAGGAS	ACAATSTCGG	GTKGGGCAAT	GTCNGGTGGG	GCAACTTTGG	120
GCTCGGRAAT	YCGGGGTAA	CGCCGGGTCT	RATGGGTSTG	GGTAATATCG	GGTTTGGTAA	180
TGCCCGCAGC	TACAATTTCG	GTTTGGCAAA	ATATGGGTGT	GGGCAATATN	GGGTTCGCTA	240
ACACCGSCAS	TGGRAATTTC	GGTATTSGGT	NACCGGTRAY	AAYCTGACCG	GGTNCGGTGG	300
TTYCAATACC	GGTAACGGGA	ATGTSGGTTS	YYYACYCCGS	GSAACGGNWW	YTTNGKTCCT	360
TMCNCTSSM	CKSAAMTSM	KMGSTSTYCT	MTYCNNGGAS	TAMTYNMCCC	CCGWAYCKSC	420
WAYCCCTCGT	CATYCCMCMC	SGSGYCCTCA	MNCCACCTTG	NGYYCCCTCC	MKMTTCYCAIT	480
CMNTCCGGTW	CCTNTMMNCC	CSCNCRYCTC	AMCNCTKSGK	CACCNATMYC	CSACKCHTCT	540
MCYMCSCAKN	MTTCCCTCN	CCTYTNNCCA	MCMCSCTCTM	TCMAACTCKC	CCGGYCKCNC	600
MYTCTCTKCC	AYNMAACCKK	TYCYWNCWYC	YMYCKCKCAG	WYKNTCTCCW	ACTCTMYNTT	660
TCTCTCNKCC	CMKACCKNTT	CTCWCSCCCC	CCACAKAYMC	YAWCMTMTCC	MCTCKACSCC	720
CYYCNMYCCM	NMCWCMTCWC	TWNAKCANCN	TTCTTCTCTC	MMYMTMACKC	WCNNTCNCCK	780
SGACCYTCTC	ACTKMKCKCK	TCTCCTTMCK	CCYMWCMTC	MKYNCCCTCC	NMTCTCKYKT	840
CCTCNMRY	CYYAKCAK	NMCTCCCAN	KMCAKCTKCT	CCCCCAKMK	ACNCKCCWC	900
CCTCCTATCC	WCTCTCWC	ATCTCKCTCW	CNYCMYMKMC	ACNCKCYAYT	CNACTMMNWN	960
CCANCTCT	CTNYCTCCK	ACCTYCKCK	CTMCKCHYMC	NRWCTYRCCT	CKKCCNCCRN	1020
CKMCMCKCTM	CTCTCCWMM	TCCWCCCAT	CTMMKSTCTC	WCNMTCCCT	CHKCCYNYNT	1080
KCYTYCCMYG	CTTCKNTCT	MCCWCCYATC	TCTMKCCTCT	CWCACYMCAC	WMTTACWNCC	1140
ACTCTCTRCW	CKCKCKMCCR	MTCTCB				1166

(2) INFORMATION FOR SEQ ID NO:329:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1230 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

11. MOLECULE TYPE: Genomic DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:329:

NGNGGNNNT	CWTACATCWN	TCTNCACCS	NGMTOWATTS	CGCGCCGCAW	NCTTGTMNAS	60
AGAATCTCAN	AAVTGGGCAC	ANATGTCTTT	TSTMTAKTGT	GGCGGGGNGC	CACGCCXTAT	120
GTGYGCTGG	GYTPACCCAA	CCCGCGGCGS	CGGGCCRACC	AGGCGGGGRA	TSCAGGCCCGC	180
GCCTGGCCGG	CCGGYTATAT	RAAGCGCCGY	TTTTKTRATA	ACGGTSCCGC	CGCCGGGTTRA	240
TTACGGGGCAA	AAVCGGKKT	TTGGGTATAT	AACCGTAATT	CCAACCAWTT	TTTTYCGGGTC	300
AAAAACVCGS	CGWGCANATC	NCGGGYCMET	RAAGCGGATT	TCGSCCAAAA	WTNTGGGGCGC	360
AAAAACCCCT	TCYTATTTTN	TGGGTATATC	GGTGGCTTGG	GCRAACGGTY	CCCGGGTTAA	420
TCCCTTCGGC	CGCGCGCCCN	AAAAACCACT	AATYCCGYTG	CGGGTGKTC	CMCAGGCGGT	480
TGCTTCGNGY	CACCTGGCCA	AAVYCCCAWT	AKATTGGGTG	CTYCKTSCCG	TTCTTGGGCT	540
CAATTACCGT	ENGGGNAAAA	RRRAAANAAA	ATGNTCTNTT	TGCTCGGYCA	YCTTTMTTGG	600
AAAAAGGGGG	ATGGTSCGGT	TYTTTACCT	CAAYCCGCNA	NEANTWACCT	YTCGSCCCCG	660
GGGGNCAAAA	CGSTTNGCTC	CGGGONAKCT	TKCTMCCCGN	ATCNAAAGGC	CNGAATTTGG	720
TTTSSSTYNA	ATTWTWKKKY	CGGWCNTTS	AAAAAAKCCA	AAASAKCCCK	YCNCAAMMYKT	780
NGGGGTYSSG	GCCKTYCTTK	SNMTTAAAC	CYCCCCAAAA	YNSGGGKKT	TCCGCYNSAT	840
KCCACNCCY	ENGGGSSGNA	AAAAAANAAY	TTTTYCCGAAA	ATCCCACTTY	TCYKTKSTRY	900
AAACCCCTT	TYMKALVTL	CTYSCNATTC	CTMTTCGAAA	TYCCGYGGCT	TNTTCCCGCK	960
CGGGGCCCC	AAWTCTCTCT	CTMCACTTCT	CTMTTCTCT	CTMCACTTCT	CTMCACTTCT	

(2) INFORMATION FOR SEQ ID NO:330:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1022 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:330:

```

NGNGGGGKNA TMAVCWTCTC ACSSGGTCTA TCGCGCGCAW CTMGTMASA GATCTCNAAY      60
TCGGGCAMNAN GCATMTCMMC CATATATAAC CATTCGCTCS GYWTSCAWCT CRAAWCTGTC      120
CTTCCKGCGG TTKTACRAAG GTGGMWTGYT CWTYCCTRAA SCCTCTCATC TCKTKTATYC      180
CTKGGGCTYC ACTTTAACSG RATKSCCTGC TTKTAYCATT RATGCAAWTA WTGGYCRAWT      240
KTTSCAGGCC PACGGCWYCT TTTTCGGGCA GRACAAATGA TTGGAWYCSC TYCGCRAGGC      300
CCGGCACCAR ACCGGGCNCC AAAGGYCCGC GCAAWTSCCT GGKTCAAAAA TGGTGCAAAAC      360
AAAMCNATCC CCGGYTTRAC CGCAGYTAMC ACAAKAAAAA TCCCWTTGGCC GCACCAWNNT      420
TTYCRATCWY CWYCCCCACC TTRAACCTSK YTGCSGTATT GCCTKCCCTGC CTCRACAGCM      480
YCNCCCKTCA AACCTGCGGT GACTCCAACT GGTCTGGYCG AASGGGGGYT CAMCGGACAA      540
AACCCCRANN TCGCCAAATT TTNCCCCGCC CYCGGGAAAN GKTGATMTTC TCSNAACCSA      600
CMGGGNNTW NAACCCGTAA CSSSGSNKGA MYNSCCSGGA ANTTTTCCCT TYNGGGCGRN      660
AAANCCTTTT AAGGTACCCC KGGNGGGGKG CCCYTTTGGG AAAACAACCC CKATTGGKTT      720
TGGAAATNTT TKNCCCCCA TTNSGGGGGG GGGCCCCCACC CMMCTTTTN TCMSCNMTYY      780
YCYTGGGAAT TNYTCCCCSG GAAYYCGGSM CCKGYCCTAA NCCCCMNWGG GKYSTGSNAR      840
GGPATMAWWT TYSTTTVYMC CCGGCNNCCC CCKAKMONT KGNTGAACMA AAKCSCGGGG      900
GSHMYMWYY YCNNGNRTT TNRGSSNMT TYMAAAMMAN GGGGKWTYY CKCCNGSCNN      960
SKTUSGGGST TTTCCNTTTS GGGSSATYKS MACCCCKTMT AYCGGGGGGT NTYTKYCCCC      1020
SC

```

(2) INFORMATION FOR SEQ ID NO:331:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1088 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:331:

```

NNCGNNKHTA TAMAYCWYCT NCACCSGGGA TFWATTGCGG CCGCAATCTT STMAASAGAT      60
CTCKAAATCG GCAMGANCCH TAWGTATTTG KOTGRASCCG ACCAGCGRGA CCTGCGCGKT      120
CKTTTCTTGC AGRGAGGCGK TGGGTGCGCG CGGTGGCAAT GCGAACCGCC CCGGAAAACN      180
CGGCAATTCY CAAAAAACAA CCGSGGGGTA KTCCTCGGCC TCCAAATMAA TAACCGTKTT      240
AAKNCAGGCH ACCGCGAAAT TTYTCTTTTC AACGNAGTNA TTTCCCGCCC NATAGGYCCG      300
TTGGGCGCTG CKTATATYCC AANTCTTAY KTNACGGGM TGGYCCMONT TCGGCTCAT      360

```

SACTGCTTTN	TTGNGGGNAM	ATANATYYTT	YCKTNGGGSK	TTCCGMTGNC	AMMAATNTCC	720
RGGGKAAMCC	AGKNTNNTCC	YTFYCCCCAA	NNTYCCYKGG	RMCTNNYYCY	TTAAANRASR	780
SAACCCCKSGG	GKCYNCNCSS	TARCCCCCAM	KAAAATTTCC	CCCSSKTTTC	TYVNNKQMRW	840
GCCCCCSAAM	ACTMTWAYTT	TCCCKCGNMN	TTTSYCKCKS	KCAMWMMWMTG	KKNCITTTTTT	900
YCSCMATAMA	CTTNGGKCCT	NTCNVSGGCG	CMAAANAAGG	CGCGSTTCTN	TTGWMAMACA	960
YNTSGNMMMA	SAAKAKWATA	AWNNTKRYK	TKNNCCCNCC	CKCKCTTSNN	TNKCCMCSKS	1020
GGGKNWNKKR	GWCTCCWCNC	CKCCCNCKNK	CKKWATMCCC	CCCCSKCCGM	NCMMNTTTKT	1080
CCC						1083

(2) INFORMATION FOR SEQ ID NO:332:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1069 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

11. MOLECULE TYPE: genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:332:

GGGGNNKYAT	MCAYCWTGTS	YACSGGGMNC	TATTGGGSCC	GCAWYTNSTM	GASAGATCTC	60
GAAYTCGGCA	MGAAAAAAGW	GATGTGCTGG	ACCTTMCSCC	GCGGGACGCR	ACCRACAAAG	120
RAASCSCGCC	ANAATATTGG	CCACAKTTGG	TCACATATTT	ACCCAATTMT	AYCAGGGGAYT	180
MCCATTSCCK	GGACCRACCG	CACAATCCCR	ATSKTGGTTT	GCRAACCCCT	ACCGTCCCCA	240
MYTYCGCCRA	STTGAACCAG	GGCRAAAAAA	CGGCCRAAWY	CTCSCCCTGA	NTCCCGCTCS	300
GSCNAATAAA	CTAGGCCCAT	TKAACGGAAU	CGGNGGCCSC	NANTTGGCCA	ACAGGTCCCT	360
ACAAAGGGGC	CCCASYCGG	CGGGWTCCTW	TTYCACNCCC	TNKTCTCKTG	CCGAATYCGG	420
WTCCPATNYC	CCWTGGGCTT	TKTGWYCKYC	KYGGGTNCCA	AWTCTNGGTA	TNCTATRGKG	480
TCCCTTAAAT	GCANATCTGG	GCKYCCATTT	NCTGGGNTTC	NATTTAMMAN	SRRCGGTTCT	540
TTCTTTCCRA	AACCCSNTGG	GCTCNMCCA	AAAAATGATN	ATAATAATGK	YGSCTTTCAA	600
ACCCCGCCCC	CCCATTCTWT	GGTTTCANE	CTCNGNGST	TAAGKTGGGA	ATTTTNTAMC	660
WENARGCCST	NATTTGGGNA	AAAACTCYKE	GGYCTCAAA	EMNYTTTTTT	GSKSSNTCCG	720
GTCTTTTCCC	CAAAACCCAA	ATTNTANGG	GGYCTCAAA	ACMCGGYCRC	RCCGGAAATT	780
TTTTTGGTTC	AACCCCAACC	TTTTCAASCC	NTTTTYYT	TCCSSCSMN	TNGSGGGGNT	840
KGGCCNTTCY	RARKKCCNMN	GGGGGWYCYN	CCCTMNTTT	CTTTTTTTT	CCGTNNMAAM	900
NGKTTCTTCA	AASMCCCCCC	GGCCCNAA	ACCCCTNAR	STTTTYCMAA	AANNWYNNGN	960
KVCCCCCCCC	MMNAAAAAAY	KCCCGGNRN	ACCMNNGGA	CCCTCCGGGN	NTTAKTTTTT	1020
TTNMGCTGCT	CTRMASYYTT	TYAMAMANER	TAMHMTTCTT	TNNPNNWNP		1080

(2) INFORMATION FOR SEQ ID NO:333:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1010 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

11. MOLECULE TYPE: genomic DNA

YCCCGGCAAA	GGGGCCACAA	CCTGCAGRGA	SCACYCRATG	GKTGYTGKTS	CNCGGGCGGG	240
CCGGKTNAAG	GGACCTGCCT	GGGKTGCSG	TMCAAAATC	WYCCGGGGT	YCGCTGGRAT	300
MCNCAGGGGT	GTCAAAAAAC	CGCAAACAGG	CACSCCANCC	NTTACGGGS	CTTAAAANGA	360
AAAAGGGCTG	ATGCCCCCAA	GGGGGCCCCG	NCCCAACCTT	CCGTTGGTCA	ACAACCCGGT	420
CTCTCKTGCC	RAATCCGRWT	CCRATNYCNC	CWTGGCCTTK	TCKYCTYCTY	CGGTACCCAA	480
ATCTGGGTAT	CCTATASTGT	CCCCTAAWTT	CCAAATCTGG	GCTGTCCATT	TSCITGGCMT	540
TCCAAATTTA	CCANCAACGG	TTTCTTNCAT	NCCAAAAACC	GNTKGGCKCC	NRACCCRAAA	600
AAATGAATAA	TAATAANNNG	KCNNTTYCNA	ACCNCCCCCC	CCCNATTCCA	IYSNGTTCCA	660
NMNCCCCCAG	NGGKTAGGTK	GGGAAANYYC	TCMACCYCA	ANCCCTWARS	TTTTNGRAAT	720
KAAACCCCTC	YCNGGGTCWW	TYMAAAAAMA	NTTATTTGGN	NGNTTTCCGG	MWNCKRKNST	780
SCCAAAATCC	MAAATANTTT	YYTGGTYCNA	TWAAAAAMCG	YGNCNMNCCC	GGAAAAWTTT	840
TINTGKTTSA	ACCCCAAAAC	YTTTTCMNA	NCSSKTTTTY	CYTTCCCCCC	AMNWTGGGYS	900
GGGNATKGYG	SCYTNTCTTA	TKTKYTYMTW	CMGGGGGGNN	MKMTICMMCCC	CCMTTTYCY	960
NYWRTTTTTN	KCCCCKTNMR	NNRAANNNGN	YTCSYNANAA	AAGCNCNCCC	SCCKNCCCN	1020
AAAAWCCCCN	NNNARAKTNT	TTMKANNRMN	SKCNKNGKY	YCCCCCCCWC	YNNMNAAAAA	1080
AATMYCCNCC	RASANMCASM	NMGGRGNRSC	CCCCCCCCST	NNNNNTMINT	TTTTTCSRA	1140
GAGCKCCSCG	MNNANMKNC	CTTTTTKCN	MNGNNGNGNN	GGNGMNCKCC	CCNAGAAMWF	1200
CTKSTCCCKS						1210

(2) INFORMATION FOR SEQ ID NO:334:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1105 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(12) SEQUENCE DESCRIPTION: SEQ ID NO:334:

NGSSSNENNA	TMCATCWYCT	SYAGSGGGMT	EWATTGCGGG	EGCAACTNGT	MAASAGATCT	60
ESAAKTCGGC	AAKANACACC	ACCGCGCTGT	MTATACACCG	FAAATTTTTT	GTKTGCCAAA	120
AGGAGAGCCG	GGGGGCCCCG	GGSYTCCAAC	TKKTTACYTR	ACCGGTCAGY	TCAGTGTTRA	180
ACCGGCTGYT	RAGGSGCCCA	CCCAACWTAA	ACGCTTTAKC	TAAGRAWYTG	GKTGGCCCCG	240
AGGCACTGTY	TGTGTYTGCC	CTCWYCGGTG	GTAGGCGCGG	TTANGGCGCG	TTGTCGCTTC	300
AMCASGSCGC	CGGTRATCCC	AKCNWTCCCC	CGGCCMRACC	GACCGGGCAC	TTTGRACGGT	360
GTGCGCAATT	CAAAAYCKYCT	GRWTCCTTCY	AAACACCACT	NAGGCGACCM	TCMSCAGCNA	420
ATMGSGRACT	TTAAGGCCCC	GGCAAAACCT	NTKACNCCCT	ACCGGCGCAA	GGTTCGCGAA	480
ATPATCCMAA	AAAANKENAT	TTGCGCGCAG	AKCAACCCAA	MMGGSTTTTG	TGCTTCCGGA	540
TTGGAAMCCA	ATTMCWGGGT	NCNNGGGAAA	AAACACCNCC	NWTAKCCMGG	TCGCGGGGGA	600
ATTTTCGFAA	GAAGCTCTNY	CCCGCTTTTT	ATNTGCTCMG	GCCTAANAAT	TCCTGGGAAT	660
AAAAAAGGTT	GGNCAAAANG	GCMAAACCCC	CACCGCACTT	NTTCCGCTTN	GGGGGGGCGN	720
TCNMSTTTAA	AWKGCCTGYT	TTSCCCAAAY	TGGGKCMAAA	NNGRKTTGGK	TTNGGCNACC	780
NTTTCCGGGC	TCGGGKGKGG	WGKYCTMMA	GGTTTNTTTT	GGCCCTKAAA	NYSGCCCGCT	840
GGGGGCCCCG	TCGGGCGGGA	NNTTTTTAMA	GKKTATCCCT	GGCCAMAAAA	ANAAGCCNYT	900
GGGGGCCCCG	TTKRWAAAMN	KUTSGCCCGG	GNNGGGCKCM	GKTTATTTMT	NNNCCSCCGG	960
TCGGGCGGAA	AAATAKMTTT	SYCCGUCGNC	CTCCGCKCNR	HKAMSMSCGC	TCCTYCTGCT	1020
TCNNTTAAAN	APGNTCTGYN	CTNNTCTGCG	NANNTNWCOD	NCCTGSCNCT	NKGCNCKNCH	1080
KAANAATATC	NGSMSTSSMY	CTTCTT				1140

- (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:335:

NGSNSNKNMN	TAMAYCWYYC	TSCACSNNGA	ACWANTGCCG	CCRMWCTNS	TMKASAGATC	60
TMGAAYTCGG	CAAGAGCGGC	AAGAGTGTGT	GCATCTGGTC	ANAGTSTMA	CRCGGTGCCG	120
CSGGTGKGR	GASCACMCAT	NTGCCGRACAC	CAAACCCCKTC	GCGGGYCACC	GGCKTCGCCT	180
GCAAAYCCT	CCAGGCCACC	TCRAACAAYW	YCTYCTGCAA	CGCARGCCGT	TYCGCGGCCG	240
RATCCTGGKT	CASYCCK	TGCCGTGCC	AAGKTACTGG	CSCAYCAAAA	CCGCTCCGGG	300
RAACRAACKT	AAWYTGCCG	AATTCNTTC	CCCTGCCCT	TGATAAATTT	NTNAAGCCAC	360
CGCAAMCCTY	CGGGCKTCTC	CTCKTGCCRA	ATYCGRWTCC	RATAYCGCCA	TGGCCTNKTC	420
KYCTYCKYCS	GTACCCAAAT	CTTGGGTATC	CTATANTKYC	CCWAAANRCA	AWTCTGGGCH	480
KTCCATKTS	TGGSKTCRA	ATTTAMMACA	NCGGTTTCTT	TCWTACCAAA	AACCSNTGGG	540
CCCCRACCRA	AAAAAGATAA	TAATAAKGTG	CCWCAAAAC	CCCCCCCCC	RRTTCAAYCG	600
GTCCARCACC	CCANGNGGTN	AGGTNGGAAT	TYTMAACCCC	CAGCCCATAA	SNITNSGNAA	660
AAACCCCCCN	GGGYMYCAAA	AMMCTTTTTC	GGGMITTCGSG	CCATKGYKCC	AAAACCAAAA	720
TMTTTCYGGT	CRWAAAAACC	GGCCCNCCCC	NAAATTTTTT	GKCAACCCCA	AACCTTTMAM	780
CCNNNTTCY	YCCNSACAA	TNGGSGGNKN	NGSSCNTTYT	TWTTTYYNNA	GGGGGRRWC	840
SNCCCNAAAN	YYCCNAANKG	NKCCCGSNMA	AAAGAGANTT	YCMKAAAAAC	CCCCNCNCCC	900
NAAAYACCCC	MAAAKWTTCM	AAASMSNNNG	YCCCCC			936

(ii) INFORMATION FOR SEQ ID NO:336:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1042 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:336:

NRNGNKNNNY	ATMMAYTCWY	YCTCCACCCG	IGNNWCWATT	CCGCGCRMW	KCTTGTMMAAS	60
AGATGCTMAA	NTCCGACAG	ASSGACAG	ASCCGCGCG	TTATYCMYCC	GYTGGTCAATC	120
ITCAACACCG	TCYTCGGG	GRATATGCG	NCSSGCGCG	GGCAACACG	YTCAAYTSC	180
TGCCCCACCG	CATATNTTAA	AAAGSTRATA	AAAGCAAAA	GGGCGCGCG	GGCTTTGGGG	240
JCGGAAACCG	GTCCCAACCT	AAAAACNCTT	TGGCACTGCG	CTGRACTTTA	AAGGTAATC	300
TKTCCCTGCT	GGCTATGCT	GGCCACAAA	CCTCTYTGCG	WGGGTCTGGC	CCTGGGYCAC	360
CCYCRCTCTT	TATMTNCTCY	CTTACACNCT	TKGSTYCAAC	CAACCCACTT	CACMAAATTS	420
TTTTTGGGKTG	GGGSSGCGCG	YGTGNNCCCH	TAATAATGCG	NTGKTGCGCC	MYCACCGGWA	480
GCATANGCTG	GGCGGCTTG	GGAAATTTCC	GAAATCATYT	GGTTCTGRAC	CCCCACAMRC	540
CTNGAAATCC	GRATCAATTC	CCCNKGGCTT	NTCYCTCTCN	GTGCCCAATY	TGGTTTCTAT	600
RKTNGCCYAA	TSCAAATGGG	TTTCURTTCG	YGGTTCCAA	TTNACAAMAS	GGTTTYTCMT	660
ACCAAAACCG	NTGGGCGGNA	TCNAAAAANA	RAAANAANKG	KCTTTYAAAC	CCCCCTCTAT	720
TCWYCGGTTC	CMRNWCCGNA	NDAANAKKIN	GAAYTTTHRA	CCCAANCCMT	ARSTTSGNAP	780
AAAACCCNCC	GGCTGCAATC	GGGCTGCAAT	GGGCTGCAAT	GGGCTGCAAT	GGGCTGCAAT	840

(2) INFORMATION FOR SEQ ID NO:337:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1073 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:337:

```

NNSGSGMKKK ATAMATCWCT CTSYACCSNG GMTCWATTGC GGCDDMAWTC TNGTMAASAG      60
ATCTCGAAYT CGGCCAAANAK ACGCMAYGTC AAGTGTRAYY CGGTACATA TCMTGCGGNG      120
TCAACMCCAA AGCCGNGTCA CCGYCTCCCT GGGGCGCCAC CCGCATCGGT RATGCAACYT      180
CGCGCGCCAC CGYCAAAAGG KTCWTTTRAGG CGCTAAAGGT CAMCAATTCC TRAGGTTCMN      240
CACCGTTNTT TGGCCCCCCC RAWTYCTRAC CGSCAATWTC GGTAAATCGGR AATTTGGGCH      300
YCGGCTTGGG CAATAAGKTN TTGGGCAACG GCGGRWTCYC NCTGGCCGGA ATTCCNCAT      360
TCCKTTAACG SKTGRACCGT TTYCCCGGYT GCGTAAYTG YTYCMTGGGC GCCYTCGGCC      420
CRNASCASY YCRCTAACGGY CMCCAGGCAA TACCKTTGSC TTTTAAACCAC CGGRATNAAY      480
TCKTACCCAC YTCALSSGTS CTGRANTTRK TNCNTGAA AANMCCACCN AACCCGGNTT      540
RATCTGCTTC MTCANCWTTT SCGGGTTTCT GCGGTTTTGR AAYCTTNATC CMTYCAAAAG      600
GTTTAMTTTC CCAANRAATT CGGYTTGCCA CTTTGGCCGS GGCTGGTTTTM CGMWCCITRR      660
AMATCCNCCS GCGGGSAAAN AMTTSGGNTT GSCCGGTCC CCGSNAATAT YCMTGGNCCT      720
GNAAATTGSS GGGATCCCN GSGNAYCCGG CCWTKGGGGK TNCCTAGTTG GWACAATTYC      780
WKCCGTTCCA AACCGGGGNC CGGGGGCTGG GSCCCTTTT CCTMYNNAAA AAGKGTITGN      840
NYTTTTTCGG CNRAANTTCA CCSKCNKINT GGNCCNAACY YYYCAANTTC CANACCTTTA      900
AASAAANCYK YGKTYVCCCG TTTTMCSSGS SANCSCCCCM NMSSKNCGGG AAAAAAGNK      960
ELNGCCTTAN CTSNKTKTTT TNKTYCCCGC NMWNNSNMCM NCBKHCNKRY NGNSNMNCCT      1020
MKYSKNNNNN SMNNNNKCSN GSNCSGMYM CMNNCNGMYK NGNKSINCCC MSC      1073

```

(2) INFORMATION FOR SEQ ID NO:338:

i. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1061 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

ii. MOLECULE TYPE: Genomic DNA

xi. SEQUENCE DESCRIPTION: SEQ ID NO:338

```

ENNGNHNWTV TMCAYCWYCT SCACSGGCTT TANTGGGGCC GCATYTNCT CKASAGATCT      60
GGATTTGGGC AMNANAARTG TCGTCTGCAA TTTTAAKKTG GTGKTDAAAY GGGTCAGGCC      120
GNSACCRACA CCTGNGTCA CCGAAAANA CAAAGCWTG AATWTCAGG GCGRAGGCGC      180
TCTCAATYCC CRASCATTA ACCGTTTGW TTTAAAGTTC CRAAGCAGG ACCCAGYTCA      240
TCCCTGGCCA AWTGGGCTG CGGTCGCTN TAACTTGAAT TTGTGAGCTT RWTCTGTGG      300
TCTTCAMCHT GGTGAAGGCC CWCCHGCHN AATACTTA TTTTCTTTC TCTTCTTTC

```

TCTNTYCGGT	GGGGCSGGCR	ANMYTTCTCT	YCCCNAAASAN	CTTAMYCCAN	TTGSSSNTCC	720
CGGKCAAWS	NGGGGGGGNA	AAGGGCCCCC	CGGNTSCKCU	GGGGKKGCCC	CYGGKTTCAA	780
AANTTTCSGG	GKTSTMSCGG	NVTSCCCCC	CSGCCAAGRA	CCGNGGTTTT	TTTTTGAACC	840
KCMANTCOSA	AMCCGCCSSC	CCCMAAAGGS	GCCTNAAWGR	RAYTTNKSCC	CNNAAACSGG	900
CCCCCAKYTY	SGGKTTTCNNC	CNCCSGKRG	CCMTSTTTMM	MRCCCTTTGN	GNKTTTTTAN	960
MGSCCTTNNC	CACCCCTYCK	GGGKCSMNNA	GAATMYWKC	CNGGGGNAN	RSCCCCCENN	1020
SGKGGGGKG	MGAGYSCCKT	CTKGCNCN	YKNTTCCCC	C		1061

(2) INFORMATION FOR SEQ ID NO:339:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 986 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

11. MOLECULE TYPE: Genomic DNA

X1. SEQUENCE DESCRIPTION: SEQ ID NO:339:

GNNGNNNKWN	ATMCAYCWYY	CTSCACCSGG	GMTCWATTGC	GSCCGCWKY	TNGTMAASAG	60
ATCTMGAAAT	CGGACANAG	CGGCACAGAG	TGTGTGCATC	TGTETCANAG	CTGTCAACGC	120
GGTGGCGGCG	GTGGTRASCA	CMCATTGCCR	AACACCAAAC	CGTCCCGCG	GYCACCGGCK	180
TGGCCTGCAA	AAVCTCCAG	GGCAGTCRA	AACAAYWYCT	CCTGCAACSC	ARSCCGSTTYC	240
GTGGCTGCRAT	CCTGJKYCAS	VTGGCKTGC	GGTGGCCCAA	GCTACTGGCS	CWYCRANACC	300
TCTYCGGGR	ACCNAACGTA	AATCTTGCCN	AATTTGCNTT	CCGCTSCCC	TTRATNAATT	360
TCTTAAAGCA	CGCAACCTY	GGGCKTCTC	CTCKTGCCRA	WTCCGRWTCC	RATNYCGCCA	420
TGGCCTNKTC	KYCTYCKYCS	GTGCCCAAT	CTTGGTATCC	TATATTGTCC	CTAAATGCAM	480
ATCTKGSTTS	TGCATNTGCT	GGGCTTCAAA	TTWAMANCAG	NGGTTCTCTY	CTTCCNAAC	540
CGSTTGGCCC	CAAAACNAAA	AATGATNATA	ATAATGGTCC	TNTCAAACCC	CGCNCCTATY	600
CNATCGSKCC	AMCCTTCRGN	GGKTANKKCG	GAATTCTTMM	AACCCCAAGC	CATAASNTTG	660
GGANAATCTY	NCNCGGGYCA	CAAAAACANY	NTTNTGGNY	SSNTTCGGMN	YCATGGSTNN	720
CAAAAACCCA	AATACTNYYS	GGYCAATAA	AAMMSGGYGC	SAMCCGGAAA	WTTTTYTTEN	780
WCAAAACCCA	AAKCTTTTTT	CNAACCCDAN	KNTYCTTNTT	CCPDMANTGG	CNSGGARTYK	840
SSCTTNCOA	ATGKYCCMAA	AGNCGGRAVA	CCARCCGAAA	TTCTTNTNTN	KNHGCTTNT	900
CTNAAAAGGG	GGTCTTCMAA	AACNCCCTCT	NCNCTCCCAA	AAKAMCCCCN	AAAGAYNTCN	960
NAANASKYCN	NNNSCCCCCC	CCMMMN				986

(2) INFORMATION FOR SEQ ID NO:340:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1074 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

11. MOLECULE TYPE: Genomic DNA

X1. SEQUENCE DESCRIPTION: SEQ ID NO:340:

ATMACCAGCY	CGCCCAGGGY	CACCTTGCCC	AAAACTCCT	GGGTGAGCCA	AATTTCGCS	360
CGGGCCAAACM	ACCANCCGCA	TYCTGGCCTC	AATCYCACCG	GGCCCGGTGY	TAAAMMANMA	420
GRATCTCKTC	MANCCCCCAN	TCAGCSYTNA	CNGCMACAGC	CGCCCTTCTT	CAMACCGCCA	480
RTACCGGGWT	CAACCGGCCS	STCAAACCTCA	ACAGGCGGNC	AGGCCTCCCC	CGGANSAAAG	540
GTCTTACSCC	NNYAANAAAA	MAAGNTCTGT	TTTCCCCCTC	CASAASNAAA	AANCCCCSGC	600
CGGGCCTTCN	NMMGGGTTTG	GGGMANANAA	AARCNC CGGN	GGAACGNATC	CGAAAMCTCC	660
CAAGTCNCMT	TWAWAACYCYN	NNAACCCCCC	ANTTTTGGGA	AAGGNTCCCC	NTTMYCCCCC	720
TTTTASGKTS	GGGMMYYCTY	TAAAAAATT	CCCCAAAAAG	CCCCGGGAG	GGTCMAMCTG	780
GGNAAATTTT	CAAMCCNWK	TTNTTYNGGT	TMCGGGGGRA	AATTTCNCTC	CCYNNNGGG	840
CSSGSNNNAT	TAYGGMSNMT	TTTNAAWTM	MSGKTSAMM	YNNKCCMNN	SNNMSMANNK	900
INAMCKCCCN	CCTCNGNGKY	CSCYNCCCSG	GNAGNGGRAS	MKCCNANMAA	AYASGNTTNK	960
CGGAAMMCNN	AATKGNNSC	CCCGASMCMN	NNNMAAATMT	CNCNKCSNN	AANRGMRACT	1020
CCCNSSNGMN	RRGAARMTNY	YCCCCGSKM	GKGNKAAAAG	GKYCCCCCM	AAAG	1074

(2) INFORMATION FOR SEQ ID NO:341:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1195 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:341:

NGNGNCHNT	MTACATCWT	CTGCACISGG	NTCWTATTC	GGCCGCAWKY	TTGTCSASAG	60
ATCTCGAAYT	CGGCCAMGAG	ACWCTGCGRA	CGCCCCCACA	NACTCTGGCG	TGTGTACCCC	120
ATTGNGGGG	TCACCGCGCC	AYTGANCCAF	TNACTGGGG	TGCCGTTCGC	CKTGCGCGGC	180
GGCTCGACGG	CKCTSCWTCT	RAAGGCWTTG	TGCACGCGAT	TGGGTTTTCT	RAACGCTGGG	240
AAAWTGGCCA	GCGCTCTGGC	TCATGGGNTC	TACGCAACGC	CNGCCCCCAA	CRCTTTCTTA	300
AATCCGGYCC	NTCCTGANC	CTTTGAAYCC	TGGGGSAAGA	ACTGGTTGCS	CNCGAVCTSC	360
TGGAACCTTK	TCNAAATCCC	GCAAKTGT	TCNTAMGYCC	CNCCGGAAGG	NGAACCTACT	420
TTCTGGWANG	TGCGCNCCK	GCGCTTATCA	CTCTGATCA	ACGCGGAAC	GGYKNNSTTG	480
KGGGAAAAAG	RRCTCAATC	MTYGGTCCCK	GGTGGKANG	CGCCCGCTGK	GYCGCNAATG	540
GAAGGCSMAG	GGTTAANGCC	MTTYCNWCCP	RSCTSTSTGA	GGKWTTCYCG	MGGANKAMN	600
MNKAMMWTTK	TCRGNSGCCW	ATGTSCCCGG	TKSTTAKAGA	ANACTYCKW	WCCGTNTVCC	660
AAAAGNTRCS	CGCMGTTTTT	CCCKMGANGN	YCTGATTTGA	GGGGGKYKCC	CCCCGGGTTC	720
TGAANKWKY	CGYAGGGGGM	NTTCCAGTCC	AGNNATNAG	AGNAAGSKTT	RYGSTSHNCC	780
TYTHGGGACT	NSCNHNSAK	ANAACNNHKT	TGCGGNTMS	AGNKTNIKGT	YCCNKTCTTC	840
TAAGAGGAGT	TATYMKCGCC	CTTGGANGMM	JAGWGMGCT	HYCCCNHKT	TCNTNSWAAA	900
TATKSAGMGG	TYCCSMAGMK	CCCGCTTTKT	TKTGANAAMN	MSMKHKKHKS	CGMGYTCTCC	960
GGGNTTGTGA	JAGTARTCCS	CCCGSMWGA	WCGGCMGNG	AGHKTTRTTS	YANTGARGCY	1020
MNNSKTMHKT	MSGSGGUNA	GGAGNGCCCC	CGANGMSTGY	NKGGNMSSNG	ARAKGATGGG	1080
GGCCNCGMNN	MGMGGANMGA	GAANGMGMR	GGGGGHTGKC	TKKCSGCGNS	CGANGPRAGAA	1140
GKTGNGSCCT	CGMGKYGHT	KTKHKTTS	YCTGNTMMN	NAGAAAAGAG	AGGGC	1195

(2) INFORMATION FOR SEQ ID NO:342:

(1) SEQUENCE CHARACTERISTICS:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:342:

CCATCTGATC	GTGGGCAACC	AGCATCGCAG	TGGGAACGAT	GCCCTCATTG	AGCATTTGCA	60
TGGTTTGTGG	AAAACCGGAG	ATGGCACTCC	AGTCGCCTTC	CCGTTCCGCT	ATCGGCTGAA	120
TTTGATTGCG	AGTGAGATAT	TTATGCCAGC	CAGCCAGACG	CAGACGCGCC	GAGACAGAAC	180
TTAATGGGCC	CGCTAACAGC	GCGATTTGCT	GGTGACCCAA	TGCGACCAGA	TGCTCCACGG	240
CCAGTCGCGT	ACCGTCCTCA	TGGGAGAAAA	TAATACTGTT	GATGGGTGTC	TGGTCAGAGA	300
CATCAAGAAA	TAACGCCCGA	ACATTAGTGC	AGGCAGCTTC	CACAGCAATG	GCATCCTGGT	360
CATCCAGCGG	ATAGTTAATG	ATCAGCCCGC	TGACGCGTTG	CGCGAGAAGA	TTGTGCACCG	420
CCGCTTTACA	GGCTTCGACG	CCGCTTCGTT	CTACCATCGA	CACCACCAGG	CTGGCACCCCA	480
GTTGATCGGC	CGGAGATTTA	ATGCGCCGSA	CAATTTGCCA	CGGCGCGTGC	AGGGCCAGAC	540
TGGAGGTGGC	AACGCCAATC	AGCAACGACT	GTTTCCCGCG	CAGTTGTTGT	GCCACCGCGT	600
TGGGAATGTA	ATTGAGCTCC	GCCATCGCGG	CTTCCACTTT	TTCCCGCGTT	TTCCGCAGAA	660
GCTGGCTGGC	CTGGTTCAAC	ACGCGGGAAA	CGGTCTGATA	AGAGACACCG	GCTACTCTGT	720
CGACATCTTA	TAACGTTACT	GTTTTCACAT	TGACCAACCT	GAATTGACTC	TCCTCCCGGC	780
CTATCTGATG	CATACCGCGA	AAGGTTTTGG	GCGATTCGAT	GTGTCCGGGG	ATCTCGACCG	840
TCTGCTTTAT	CGGACTCCTG	CATTAGGAAG	CAGTCCAGTA	GTAGGTTGAG	GCCTGTGAGC	900
ACCGCCCGCG	TAAGGAATGG	TGCATGCAAG	GAGATGGCGC	CCAACAGTCC	CCCGGCCACG	960
GGGCTTGCCA	CGATACCCAG	GCGGAAACAA	GCGTTCATGA	GCCCGAAGTG	GCGAGCCCGA	1020
TCTTCCCAT	CGGTGATGTC	CGCGATATAG	GCGCCAGCAA	CCGCACCTGT	GCGCCCGGTT	1080
ATGCGCGCCA	CGATGCGTCC	GCGCTAGAGG	ATCGAGATCT	CGATCCCGCT	AAATTAAATC	1140
GACTCACTAT	AGGGGAATTG	TGAGCGGATA	ACAATTCCCG	TGATGCAATA	ATTTTGTCTA	1200
ACTTTAAGAA	GGAGATATAC	ATATGGGCCA	TCATCATCAT	CATCAGGTGA	TGACATCTAT	1260
CGGACCCAGC	CCCACATCCG	GCGAACACGG	GCGCGCGGAG	GCGGTCCAGC	GGGCGCGGGA	1320
TAGGCTCGAT	GACATCCGCG	TGCGTCGGGT	CATTGAGCAG	GACATGGCCG	TGGACAGCCG	1380
TAGCCAGATG	ACCTACCGCA	TCAAGCTCGA	AGTGTCTGTC	AAGATGAGGC	CGGCGCAACG	1440
GAAGGCTCG	AAACGACCGA	CGGTTTCCCG	TGAAACCGGC	GCCGGCGCGG	GTACTCTCGC	1500
GAATACCCCG	CGGTCTGCGC	CGGTGACCTT	GCGCGAGAGC	GGTAGCACCG	TGCTCTACCG	1560
CGTCTTCAAC	CTGTGGGGTC	CGGCTTTTCA	CGAGAGGTAT	CGGACAGTCA	TGACATCCCG	1620
TGAGGCGACC	CGTTCTGGTG	CGCGGATCCG	CGAGGCGCCG	CGCGGACCGG	TGACATCCCG	1680
CGCTCCGAC	CGCTATCTGT	CGGAAGGTGA	TATGCGCGCG	CACAGGGGCG	TGATGAACAT	1740
CGCTCTAGCG	ATCTCTCGCT	AGCAGGTGAA	GTACCAAGTG	CGCGAGTGA	CGGAGACCTT	1800
CGCTCTGAAG	CGAAAGGTGC	TGCGCGCCAT	GTACCAAGGG	ACCATCAAAA	CGTGGGACGA	1860
CGCTCGAGAT	CTCTGGCTCA	ACCGCGCGCT	GAACCTCCCG	CGCACCGCGG	TAGTTCCCGT	1920
CGACGCTCCG	GACGGGTCCG	GTGACAGCTT	CTTCTTCCCG	CAGTACCTGT	CGAGGCAAGA	1980
TGCGCAAGGC	TGGGGCAAGT	CGCGCGCGCT	CGCGACCAAC	CTGCACTTCC	CGCGCGTCCG	2040
CGCTCGCGTG	CGTGAGAACC	GTAAATGCGG	ATGCTGAGAG	CGTCTGCGCG	AGACATCCCG	2100
CGCTCTCGCT	TATATGCGCA	TGAGCTTCTT	TGAGGAGGCG	ATGCAACCGG	TACTGCGCGA	2160
CGCTCAACTA	CGCAATAGCT	ATGCAATGTT	TTGCTTCTCG	ATGCGCGAAA	CGATTCTAGT	2220
CGCTCGCGCT	CGCTTCTCAT	TGAAAAATCT	TGAGGAGGCG	CGGATTTGGA	TGATCGAGCT	2280
CGCTCGCGCG	JAGGCTTACG	CGATCATGAA	TGAGGAGGCG	CGATCTGCTA	AGAACCGGGA	2340
CGCTCGCGCG	CGCGCGCGCG	AGAGCTTCCA	AGATTTCTTG	CGTGGGGCGA	TGACCGAGCT	2400
CGCTCGCGCG	TGCTTCTCGG	ACGAGGTTCA	TTTCTAGCGG	CTCGCGCGCG	CGGTCTGAA	2460
CGCTCTGAGG	CGCTTCTCGG	CGAGCTTCTG	CGCTTCTGAG	ATGAGACCGG	ATGCGCTTAC	2520
CGCTCTGAGG	JAGGCGAGGA	ATTTCTGAGG	CGCTTCTGAG	CGCTGAGAAA	CGGAGATCGA	2580
CGCTCTGAGG	TGAGCGCGAG	CTTCTTCTCA	CGCTTCTGAG	CGCGCGCGCG	CGGCGAGCTT	2640
CGCTCTGAGG	CGCTTCTCTG	CTTCTCTGAA	CGCTTCTGAG	CGCTTCTGAG	CGGCGAGCTT	2700
CGCTCTGAGG	CGCTTCTCTG	CTTCTCTGAA	CGCTTCTGAG	CGCTTCTGAG	CGGCGAGCTT	2760

```

AACCGTCTCG CTYGACGCCA ACGGGGTGTC TGAAGCGCG TCGTATTACG AAGTCAAGTT 3120
CAGCGATCCG AGTAAGCCGA ACGGCCAGAT CTGGACGGGC GTAATCGGCT CGCCCGCGGC 3180
GAACGCACCG GACGCCGGGC CCCCTCAGCG CTGGTTTGTG GTATGGCTCG GGACCGCCAA 3240
CAACCCGGTG GACAAGGGCG CGGCCAAGGC GCTGGCCGAA TCGATCCGGC CTTTGGTCTG 3300
CCCGCCGGCG GCGCCGGGCG GGGAAAGTCG TCCTACCCCG ACGACACCGA CACCGCAGCG 3360
GACCTTACCG GCCTGAGAAT TCTGCAGATA TCCATCACAC TGGCGGGCCG TCGAGCACCA 3420
CCACCACCAC CACTGAGATC CGGCTGCTAA CAAAGCCCCG AAGGAAGCTG AGTTGGCTGC 3480
TGCCACCGCT GAGCAATAAC TAGCATAACC CCTTGGGGCC TCTAAACGGG TCTTGAGGGG 3540
TTTITTGCTG AAAGGAGGAA CTATATCCGG AT 3572

```

(2) INFORMATION FOR SEQ ID NO:343:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:343:

```

Val Gln Phe Gln Ser Gly Gly Asp Asn Ser Pro Ala Val Tyr Xaa Xaa
 1             5             10             15
Asp Gly Xaa Arg
                20

```

(2) INFORMATION FOR SEQ ID NO:344:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:344:

```

Thr Thr Val Phe Val Val Thr His Ala Arg
 1             10

```

(2) INFORMATION FOR SEQ ID NO:345:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

1 S 10

(2) INFORMATION FOR SEQ ID NO: 346:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:346:

Asp Ala Gly Lys Xaa Ala Gly Xaa Asp Val Xaa Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO:347:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:347:

Thr Xaa Glu Glu Xaa Glu Glu Ser Phe Asn Ser Ala Ala Pro Gly Asn
1 5 10 15
Xaa Lys

2 INFORMATION FOR SEQ ID NO: 348:

1. SEQUENCE CHARACTERISTICS

- A. LENGTH: 27 base pairs
B. TYPE: nucleic acid
C. STRANDEDNESS: single
D. TOPOLOGY: linear

11 MOLECULE TYPE: Other

X21 SEQUENCE DESCRIPTION: SEQ ID NO. 348.

TTAGTTAGTA CTCAGTCCCA JACCGTG

THE INFORMATION REPORT SER. NO. 24

[illegible]

GCAGTGACGA ATTCACCTCG ACTCC

25

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2412 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

XL SEQUENCE DESCRIPTION: SEQ ID NO:350:

CATATGGGGCC	ATCATCATCA	TCATCACGTG	ATCGACATCA	TGGGGACCAG	CCCCACATCC	60
TGGGAACAGG	CGGCGGCGGA	GGGGGTCCAG	CGGCGCGGGG	ATAGCOTCGA	TGACATCCGC	120
GTGGGTGGG	TCATTGAGCA	GGACATGGCC	GTGGACAGCG	CCGCGAAGAT	CACCTACCGC	180
ATCAAATCCG	AAGTGTCTGT	CAAGATGAGG	CTGGGCGCAAC	CGAGGGGCTC	GAACACACC	240
AGCGGTTCG	CTGAAACGGG	CGCGGGCGCC	GCTACTGTCC	CGACTACCC	CGCGTCTCTC	300
CGCGGTACGT	TGGCGGAGAC	CGGTAGCAAC	CTGCTCTACC	CGCTCTTCAA	CTGTGTGGGT	360
CGCGGTTTTG	ACGAGAGGTA	TCGGAACTC	ATGATCAAGG	CTCAGGGCAG	CGGTTCTGGT	420
CGCGGATCCG	CGCAGGCGCG	CGCGGGAACG	CTCAAAATTC	GGCGCTCCGA	CGCTATCTG	480
TCGGAAAGTG	ATATGGCCCG	GCACAAGGGG	CTGATCAACA	TGGCGCTAGC	CATCTCGCT	540
CAGCAAGTCA	ACTACAACCT	CGCGGGAGTG	AGCGAGCAAC	TCAAGCTGAA	CGGAAAAGTG	600
CTCGCTGGCA	TGTACCAAGG	TACCATCAAA	ACCTGGGAGC	ACCTCGAGAT	TGCTGCTCT	660
AACCGTGGCG	TGAACCTGCC	CGGCACCGCG	GTAGTTCTCT	TGCAACGGTG	CGACGGGTCC	720
GCTGACACCT	TCTTGTTCAC	CGAGTACCTG	TGCAAGCAAG	ATCGGTAGGG	TGGGGCAAG	780
TGCGCTGGGT	TCGGCACCA	CTCTGACTTC	CGCGCGGTTC	CGGGTGGCTT	TGGTGAGAAC	840
CGCAACGGCG	CGATCTGTAC	CGCTTGGCTC	JAGACAGGGG	TCTGGCTTCT	CTATATGGC	900
ATCACTTTCC	TGACCAAGGC	CGATCAACCG	CGACTCGGG	AGGCGCAACT	AGGCAATAGC	960
CTCGGCAATT	TCTTGTTCCT	CGACGCGCAA	AGCATTCAGG	CGCGCGGCGT	TGGCTTGGCA	1020
TGGAAAACCG	CGCGCAACCA	GGCGATTTCT	ATGATCGAGC	GGCGCGCTCC	GGACCGGTAC	1080
CGCATCATCA	ACTACAGTGA	CGCATCTCTC	AACAAACCGC	AAAAAGACCT	CGCCAGCGCG	1140
TAGACTTTGC	AGGCATTTCT	CGATCTGGCG	ATCAACGATC	CGAAAGAGCG	CTGGTTCTCT	1200
JAGCAAGTTT	ATTTCAAGCG	GCTCGCGCTC	CGCTCTCTCA	AGTTCTCTCA	CGCTTCTATC	1260
CGCAAGATT	CGAGCTCTCA	CATCAAGACG	TATCGCGCTA	CGCTCTCTCA	CGAGGCGAGG	1320
AGTTCTGAGG	CGATCTCTCG	CGACTCGAAA	ATCGAGATCG	ATAGGTCTCA	CTCGAGCGCA	1380
CGTTCTTTCT	AGGCGAGCTG	CGCGCGCGCG	CGCGCGAGCG	CGCGCGAGCG	CGCGGTGGTG	1440
CGTTCTTACG	AAGCAAGCAA	TAGCAGAAAG	CAGGAACTCT	AGCAATCTCT	GAGCAATATT	1500
CGTCAAGCTG	CGCTCGAATA	CTCGAGGGCT	GATCAGGAGC	AGCAAGAGCT	GCTCTCTCTC	1560
CAAAATGGCT	TTCTGCGCAC	AACGCGCGCG	TGCGCGCTCT	CGACTCTCTC	AGCGCCACCC	1620
CGACGGGCGA	CAGCTGTCTC	CGCGCGACCA	CGCGCGCGCG	CGAACAGCGC	GAATCGCCAG	1680
CGGTGCGATC	CGAACGCGAG	ACCTCGCGCG	CGCGACCGCA	AGCGACCGCG	GCACCTCTCT	1740
ATTTCTTACG	AGCAAGCGCA	ACCTCTCTCT	ATCGACAGCT	CGCTCTGAGG	ATTCAGCTTT	1800
CGCTCTTCTG	TGCTGTGGGT	CGAGCTTCTC	CGCGCGACT	TGACTACCG	TCAGGCACTC	1860

```

ATCGGCTCGC CCGCGGCGAA CGCACCGGAC GCGGGGCCCC CTCAGCGCTG GTTTGTGGTA 2220
TGGCTCGGGA CCGCCAACAA CCGGVTGGAC AAGGGCGCGG CCAAGGCGCT GGCCGAATCG 2280
ATCCGGCCTT TGGTCGCCCC GCGCGCGGCG CCGGCACCGG CTCCTGCAGA GCGCGCTCCG 2340
GCGCCGCGCG CCGCCGGGGA AGTCGCTCCT ACCCCGACGA CACCGACACC GCAGCGGACC 2400
TTACCGGCCT GA 2412

```

(2) INFORMATION FOR SEQ ID NO:351:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 802 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:351:

```

Met Gly His His His His His His Val Ile Asp Ile Ile Gly Thr Ser
 1           5           10           15
Pro Thr Ser Trp Glu Gln Ala Ala Ala Glu Ala Val Gln Arg Ala Arg
 20           25           30
Asp Ser Val Asp Asp Ile Arg Val Ala Arg Val Ile Glu Gln Asp Met
 35           40           45
Ala Val Asp Ser Ala Gly Lys Ile Thr Tyr Arg Ile Lys Leu Glu Val
 50           55           60
Ser Phe Lys Met Arg Pro Ala Gln Pro Arg Gly Ser Lys Pro Pro Ser
 65           70           75           80
Gly Ser Pro Glu Thr Gly Ala Gly Ala Gly Thr Val Ala Thr Thr Pro
 85           90           95
Ala Ser Ser Pro Val Thr Leu Ala Gln Thr Gly Ser Thr Leu Leu Tyr
100           105           110
Pro Leu Phe Asn Leu Trp Gly Pro Ala Phe His Glu Arg Tyr Pro Asn
115           120           125
Val Thr Ile Thr Ala Gln Gly Thr Gly Ser Gly Ala Gly Ile Ala Gln
130           135           140
Ala Ala Ala Gly Thr Val Asn Ile Gly Ala Ser Asp Ala Tyr Leu Ser
145           150           155           160
His Gly Asp Met Ala Ala His Lys Gly Leu Met Asn Ile Ala Leu Ala
165           170           175
Ile Ser Ala His His Val Asn Tyr Asn Leu Pro Gly Val Ser Gln His
180           185           190
Leu Lys Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile
195           200           205
Lys Thr Thr Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn
210           215           220
Leu Pro Gly Thr Ala Val Val Pro Leu His Arg Ser Asn Gly Ser Gly
225           230           235           240
Asp Thr Phe Leu Pro Thr Gln Tyr Leu Ser Leu His Asn Pro Gln Gly
245           250           255

```

272

290 295 300
 Gln Ala Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser
 305 310 315 320
 Gly Asn Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala
 325 330 335
 Gly Phe Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp
 340 345 350
 Gly Pro Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile
 355 360 365
 Val Asn Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala
 370 375 380
 Phe Leu His Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp
 385 390 395 400
 Gln Val His Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp
 405 410 415
 Ala Leu Ile Ala Thr Ile Ser Ser Ala Glu Met Lys Thr Asp Ala Ala
 420 425 430
 Thr Leu Ala Gln Glu Ala Gly Asn Phe Glu Arg Ile Ser Gly Asp Leu
 435 440 445
 Lys Thr Gln Ile Asp Gln Val Glu Ser Thr Ala Gly Ser Leu Gln Gly
 450 455 460
 Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln Ala Ala Val Val Arg
 465 470 475 480
 Phe Gln Glu Ala Ala Asn Lys Gln Lys Gln Glu Leu Asp Glu Ile Ser
 485 490 495
 Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg Ala Asp Glu Glu
 500 505 510
 Gln Gln Gln Ala Leu Ser Ser Gln Met Gly Phe Val Pro Thr Thr Ala
 515 520 525
 Ala Ser Pro Pro Ser Thr Ala Ala Ala Pro Pro Ala Pro Ala Thr Pro
 530 535 540
 Val Ala Pro Pro Pro Pro Ala Ala Ala Asn Thr Pro Asn Ala Gln Pro
 545 550 555 560
 Gly Asp Pro Asn Ala Ala Pro Pro Pro Ala Asp Pro Asn Ala Pro Pro
 565 570 575
 Pro Pro Val Ile Ala Pro Asn Ala Pro Gln Pro Val Arg Ile Asp Asn
 580 585 590
 Pro Val Gly Gly Phe Ser Phe Ala Leu Pro Ala Gly Trp Val Glu Ser
 595 600 605
 Asn Ala Ala His Phe Asp Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr
 610 615 620
 Gly Asp Pro His Phe Leu Gly Gln Pro Pro Pro Val Ala Asn Asp Thr
 625 630 635 640
 Arg Ile Val Leu Gly Arg Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu
 645 650 655
 Ala Thr Asp Ser Lys Ala Ala Ala Arg Leu Gly Ser Asp Met Gly Glu
 660 665 670
 Phe Tyr Met Pro Tyr Pro Gly Thr Arg Ile Asn Gln Glu Thr Val Ser
 675 680 685
 Leu Asn Ala Asn Gly Val Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys

Phe	Val	Val	Trp	Leu	Gly	Thr	Ala	Asn	Asn	Pro	Val	Asp	Lys	Gly	Ala
				740					745					750	
Ala	Lys	Ala	Leu	Ala	Glu	Ser	Ile	Arg	Pro	Leu	Val	Ala	Pro	Pro	Pro
		755					760					765			
Ala	Pro	Ala	Pro	Ala	Pro	Ala	Glu	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala
	770					775					780				
Gly	Glu	Val	Ala	Pro	Thr	Pro	Thr	Thr	Pro	Thr	Pro	Gln	Arg	Thr	Leu
785					790					795					800
Pro	Ala														

(2) INFORMATION FOR SEQ ID NO:352:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Other

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:352:

GGATCCAAAC CACCGAGCGG TTCGCCTGAA ACGG

34

(2) INFORMATION FOR SEQ ID NO:353:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Other

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:353:

CGCTGCGAAT TCACCTCCGG AGGAAATCGT CCGCATC

37

(2) INFORMATION FOR SEQ ID NO:354:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1962 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:354:

```

GCCGCCGGGA CGGTCAACAT TGGGGCCTTC GACGCCTATC TGTCGGGAAGG TGATATGGCC 300
GCGCACAAAG GGCTGATGAA CATCGCCTTA GCCATCTCCG CTCAGCAGGT CAACTACAA 360
CTGCCCGGAG TGAGCGAGCA CCTCAAGCTG AACGGAAAAG TCCTGGCGGC CATGTACCAG 420
GGCACCATCA AAACCTGGGA CGACCCCGAG ATCGCTGCGC TCAACCCCGG CGTGAACCTG 480
CCCGGCACCG CGGTAGTTCC GCTGCACCGC TCCGACGGGT CCGGTGACAC CTTCTTGTTT 540
ACCCAGTACC TGTCCAAGCA AGATCCCGAG GGCTGGGGCA AGTCGCCCGG CTTCCGGACC 600
ACCGTCCGACT TCCCGGCGGT GCCGGGTGCG CTGGGTGAGA ACGGCAACGG CGGCATGGTG 660
ACCGGTTGCG CCGAGACACC GGGCTGCGTG GCCTATATCG GCATCAGCTT CCTCGACCAG 720
GCCAGTCAAA GGGGACTCGG CGAGGCCCAA CTAGGCAATA GCTCTGGCAA TTTCTTGTTG 780
CCCGACGCGC AAAGCATTCA GGCCGCGGCG GCTGGCTTCG CATCGAAAAC CCCGCGAAC 840
CAGGCGATTT CGATGATCGA CGGGCCCGCC CCGGACGGCT ACCCGATCAT CAACTACGAG 900
TACGCCATCG TCAACAACCG GCAAAGGAC GCCGCCACCG CGCAGACCTT GCAGGCATTT 960
CTGCACTGGG CGATCACCAG CGGCAACAAG GCCTCGTTCC TCGACCAGGT TCATTTCAG 1020
CCGCTGCCCG CCGCGGTGGT GAAGTTGTCT GACGCGTTGA TCGCGACGAT TTCTTCGGGA 1080
GGTGGCAGTG GGGGAGGCTC AGGTGGAGGT TCTGGCGGGA CGGTGCCCCA AACGGCCGCC 1140
TCGCCGCCGT CGACCGCTGC AGCGCCACCC GCACCGGCGA CACCTGTTGC CCCCCACCA 1200
CGGGCCGCCG CCGACCGCC GAATGCGAG CCGGGCGATC CCAACGCAGC ACCTCGGCCG 1260
GCCGACCGCA ACGCACCGCC GCCACCTCTC ATTGCCCGCA ACGCACCGCA ACCTGTCGGG 1320
ATCGACAACC CGGTGGAGG ATTCAGCTTC GCGCTGCTTG CTGGCTGGGT GGAGTCTGAC 1380
GCCGCCACT TCGACTACCG TTCAGCACTC CTCAGCAAAA CCACCGGGGA CCCGCCATTT 1440
CCCGACAGC CGCCGCCGT GCGCAATGAC ACCCGTATCG TGCTCGGCCG GCTAGACCAA 1500
AAGCTTTAGC CCAGCGCGCA AGCCACCGAC TCCAAGGCCG CGGCCCGGTT GGCTCGGAC 1560
ATGGTGTAGT TCTATATGCC CTACCCGGGT ACCCGGATCA ACCAGGAAAC CGTCTCGCTC 1620
GACGCAACG GGTGTCTGG AAGCGCTCG TATTACGAAG TCAAGTTTCA CGATCCGAGT 1680
AAACCGAAG GCCAGATCTG GACGGCGCTA ATCGGCTCGC CCGCGGCGAA CGCACCGGAC 1740
GCCGGCCCGC CTCAGCGCTG TTTTGTGCTA TGGCTCGGGA CCGCCAACAA CCGGTGGAC 1800
AAGGCGCGCG CCAAGGCGCT GCGCAATCG ATCGCGCTT TGGTCCCGCC GCGCCCGCG 1860
CCCGIACCG CTCCTGCAGA GCGCGCTCG GCGCGGCGCG CGGCCGGGA AGTCGCTCCT 1920
ACCGIACCG CACCGACACC GCAGCGGACT TTACCGGCT GA 1960

```

2. INFORMATION FOR SEQ ID NO:55:

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 552 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

11. MOLECULE TYPE: protein

11. SEQUENCE DESCRIPTION: SEQ ID NO:55

```

Met His His Ala His His His His Gly Ser Gly Pro Pro Ser Gly Ser
1 7 13 19 25 31 37 43 49 55 61 67 73 79 85
Pro Val Thr Gly Ala Gly Ala Gly Thr Val Ala Thr Thr Pro Ala Ser
91 97 103 109 115 121 127 133 139 145 151 157 163 169 175
Ser Thr Val Thr Leu Ala Gly Thr Gly Ser Thr Leu Leu Tyr Pro Leu
181 187 193 199 205 211 217 223 229 235 241 247 253 259 265
Pro Leu Leu Tyr Gly Pro Ala Pro His His Ala Thr Pro Arg Ala Thr
271 277 283 289 295 301 307 313 319 325 331 337 343 349 355
Thr Thr Ala His His Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
361 367 373 379 385 391 397 403 409 415 421 427 433 439 445

```


275

100 105 110
Ala Gln Gln Val Asn Tyr Asn Leu Pro Gly Val Ser Glu His Leu Lys
115 120 125
Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile Lys Thr
130 135 140
Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn Leu Pro
145 150 155 160
Gly Thr Ala Val Val Pro Leu His Arg Ser Asp Gly Ser Gly Asp Thr
165 170 175
Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Glu Gly Trp Gly
180 185 190
Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val Pro Gly
195 200 205
Ala Leu Gly Glu Asn Gly Asn Gly Gly Met Val Thr Gly Cys Ala Glu
210 215 220
Thr Pro Gly Cys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp Gln Ala
225 230 235 240
Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser Gly Asn
245 250 255
Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala Gly Phe
260 265 270
Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp Gly Pro
275 280 285
Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile Val Asn
290 295 300
Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala Phe Leu
305 310 315 320
His Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp Gln Val
325 330 335
His Phe Gln Pro Leu Pro Pro Ala Val Lys Leu Ser Asp Ala Leu
340 345 350
Ile Ala Thr Ile Ser Ser Gly Gly Gly Ser Gly Gly Ser Gly Gly
355 360 365
Gly Ser Gly Gly Ser Val Pro Thr Thr Ala Ala Ser Pro Pro Ser Thr
370 375 380
Ala Ala Ala Pro Pro Ala Pro Ala Thr Pro Val Ala Pro Pro Pro Pro
385 390 395 400
Ala Ala Ala Asn Thr Pro Asn Ala Gln Pro Gly Asp Pro Asn Ala Ala
405 410 415
Pro Pro Pro Ala Asn Pro Asn Ala Pro Pro Val Pro Val Ile Ala Pro
420 425 430
Asn Ala Pro Gln Thr Val Arg Ile Asn Asn Pro Val Gly Ile Phe Leu
435 440 445
Pro Ala Leu Pro Ala Gln Thr Val Gln Ser Ala Ala Ala His Phe Asp
450 455 460
Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr Gly Asp Pro Pro Phe Pro
465 470 475 480
Gly Gln Pro Pro Pro Val Ala Asn Asn Thr Arg Ile Val Leu Gly Arg
485 490 495
Leu Asn Gln Lys Leu Thr Ala Ser Ala Ile Ala Thr Asp Ser Lys Ala
500

Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys Phe Ser Asp Pro Ser Lys
 545 550 555 560
 Pro Asn Gly Gln Ile Trp Thr Gly Val Ile Gly Ser Pro Ala Ala Asn
 565 570 575
 Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp Phe Val Val Trp Leu Gly
 580 585 590
 Thr Ala Asn Asn Pro Val Asp Lys Gly Ala Ala Lys Ala Leu Ala Glu
 595 600 605
 Ser Ile Arg Pro Leu Val Ala Pro Pro Pro Ala Pro Ala Pro Ala Pro
 610 615 620
 Ala Glu Pro Ala Pro Ala Pro Ala Pro Ala Gly Glu Val Ala Pro Thr
 625 630 635 640
 Pro Thr Thr Pro Thr Pro Gln Arg Thr Leu Pro Ala
 645 650

CLAIMS

1. A polypeptide comprising an immunogenic portion of a soluble *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen has an N-terminal sequence selected from the group consisting of:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu; (SEQ ID No. 120)
- (b) Ala-Val-Glu-Ser-Gly-Met-Ile-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser; (SEQ ID No. 121)
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg; (SEQ ID No. 122)
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro; (SEQ ID No. 123)
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val; (SEQ ID No. 124)
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro; (SEQ ID No. 125)
- (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Thr-Ala-Ala-Ser-Pro-Pro-Ser; (SEQ ID No. 126)
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly; (SEQ ID No. 127)
- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Ileu-Thr-Ser-Ileu-Ileu-Asn-Ser-Ileu-Ala-Asn-Pro-Asn-Val-Ser-Phe-Xaa-Asn; (SEQ ID No. 128) and
- (j) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Ileu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 130)

wherein Xaa may be any amino acid

substitutions and/or modifications, wherein said antigen has an N-terminal sequence selected from the group consisting of:

- (a) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID No. 129) and
- (b) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID No. 137), wherein Xaa may be any amino acid.

3. A polypeptide comprising an immunogenic portion of a soluble *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID Nos.: 1, 2, 4-10, 13-25, 52, 99 and 101, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 1, 2, 4-10, 13-25, 52, 99 and 101 or a complement thereof under moderately stringent conditions.

4. A polypeptide comprising an immunogenic portion of a *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID Nos.: 26-51, 138, 139, 163-183, 201, 240, 242-247, 253-256, 295-298, 309, 316, 318-320, 322, 324, 328, 329, 333, 335, 337, 339 and 341, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 26-51, 138, 139, 163-183, 201, 240, 242-247, 253-256, 295-298, 309, 316, 318-320, 322, 324, 328, 329, 333, 335, 337, 339 and 341 or a complement thereof under moderately stringent conditions.

5. A DNA molecule comprising a nucleotide sequence encoding a polypeptide according to any one of claims 1-4.

7. A host cell transformed with an expression vector according to claim 6.
8. The host cell of claim 7 wherein the host cell is selected from the group consisting of *E. coli*, yeast and mammalian cells.
9. A pharmaceutical composition comprising one or more polypeptides according to any one of claims 1-4 and a physiologically acceptable carrier.
10. A pharmaceutical composition comprising one or more DNA molecules according to claim 5 and a physiologically acceptable carrier.
11. A pharmaceutical composition comprising one or more DNA sequences recited in SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347; and a physiologically acceptable carrier.
12. A vaccine comprising one or more polypeptides according to any one of claims 1-4 and a non-specific immune response enhancer.
13. A vaccine comprising
a polypeptide having an N-terminal sequence selected from the group consisting of sequences recited in SEQ ID NOs: 134 and 135, and
a non-specific immune response enhancer
14. A vaccine comprising
one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347, the complements of said sequences, and DNA sequences that

a non-specific immune response enhancer.

15. The vaccine of claims 12-14 wherein the non-specific immune response enhancer is an adjuvant.

16. A vaccine comprising one or more DNA molecules according to claim 5 and a non-specific immune response enhancer.

17. A vaccine comprising one or more DNA sequences recited in SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215, 225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347; and a non-specific immune response enhancer.

18. The vaccine of claims 16 or 17 wherein the non-specific immune response enhancer is an adjuvant.

19. A method for inducing protective immunity in a patient, comprising administering to a patient a pharmaceutical composition according to any one of claims 9-11.

20. A method for inducing protective immunity in a patient, comprising administering to a patient a vaccine according to any one of claims 12-18.

21. A fusion protein comprising two or more polypeptides according to any one of claims 1-4.

22. A fusion protein comprising one or more polypeptides according to any one of claims 1-4 and ESAT-6.

23. A fusion protein comprising one or more polypeptides according to any one of claims 1-4 and the *M. tuberculosis* antigen 18 kDa secretory protein.

24. A pharmaceutical composition comprising a fusion protein according to any one of claims 21-23 and a physiologically acceptable carrier.

25. A vaccine comprising a fusion protein according to any one of claims 21-23 and a non-specific immune response enhancer.

26. The vaccine of claim 25 wherein the non-specific immune response enhancer is an adjuvant.

27. A method for inducing protective immunity in a patient, comprising administering to a patient a pharmaceutical composition according to claim 24.

28. A method for inducing protective immunity in a patient, comprising administering to a patient a vaccine according to claims 25 or 26.

29. A method for detecting tuberculosis in a patient, comprising:

(a) contacting dermal cells of a patient with one or more polypeptides according to any one of claims 1-4; and

(b) detecting an immune response on the patient's skin and therefrom detecting tuberculosis in the patient.

30. A method for detecting tuberculosis in a patient, comprising:

(a) contacting dermal cells of a patient with a polypeptide having an N-terminal sequence selected from the group consisting of sequences recited in SEQ ID NO: 134 and 135; and

(b) detecting an immune response on the patient's skin and therefrom detecting tuberculosis in the patient.

31. A method for detecting tuberculosis in a patient, comprising:

310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347; and

(b) detecting an immune response on the patient's skin and therefrom detecting tuberculosis in the patient.

32. The method of any one of claims 29-31 wherein the immune response is induration.

33. A diagnostic kit comprising:

(a) a polypeptide according to any one of claims 1-4; and
(b) apparatus sufficient to contact said polypeptide with the dermal cells of a patient.

34. A diagnostic kit comprising:

(a) a polypeptide having an N-terminal sequence selected from the group consisting of sequences recited in SEQ ID NO: 134 and 135; and
(b) apparatus sufficient to contact said polypeptide with the dermal cells of a patient.

35. A diagnostic kit comprising:

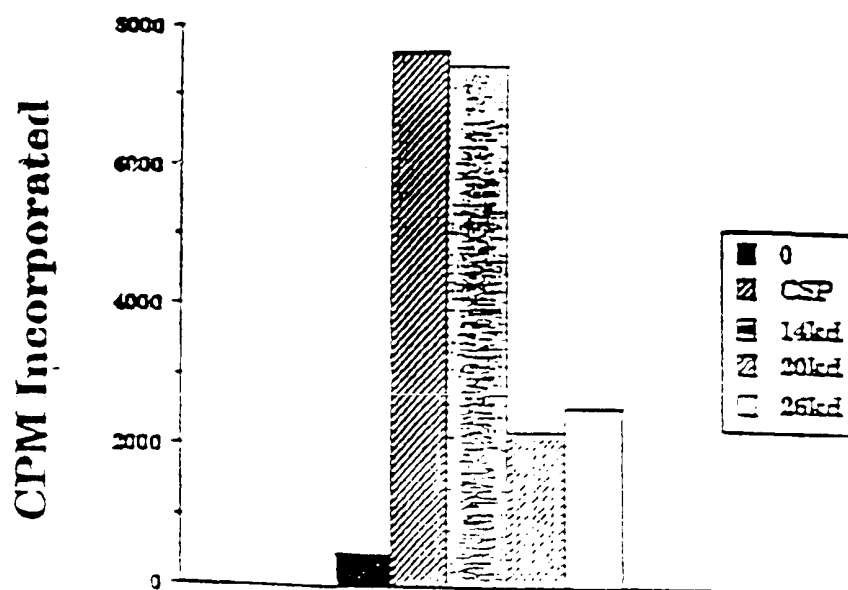
(a) a polypeptide encoded by a DNA sequence selected from the group consisting of SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347; and

36. A diagnostic kit comprising:

- (a) a fusion protein according to any one of claims 21-23; and
- (b) apparatus sufficient to contact said fusion protein with the dermal cells of a patient.

37. A fusion protein according to claim 23 comprising an amino acid sequence selected from the group consisting of sequences recited in SEQ ID NO: 153, 209, 351 and 355.

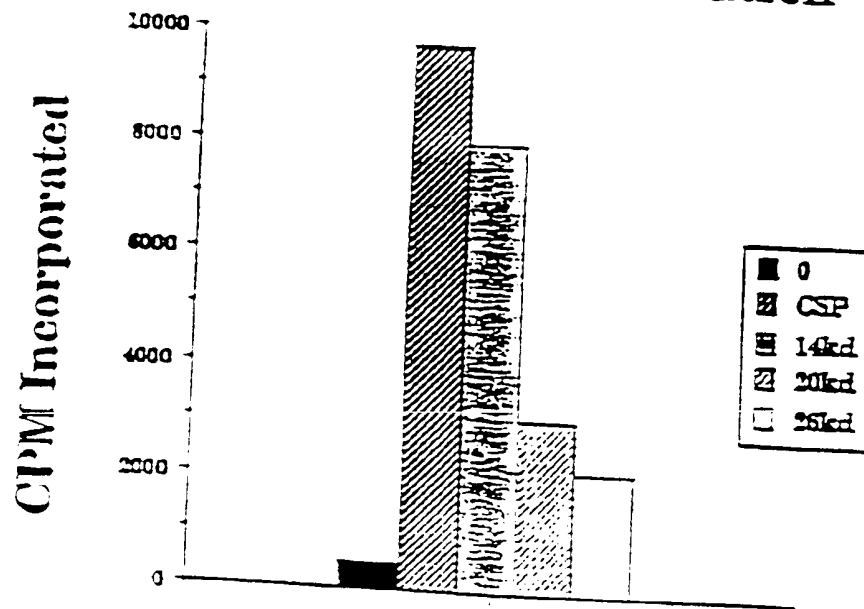
D7 T Cell Proliferation

D7 IFN_g

O.D. 450-570

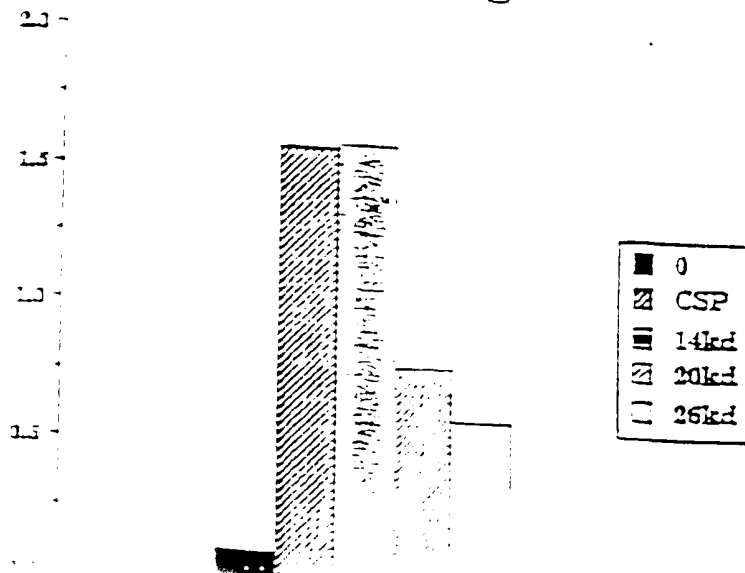


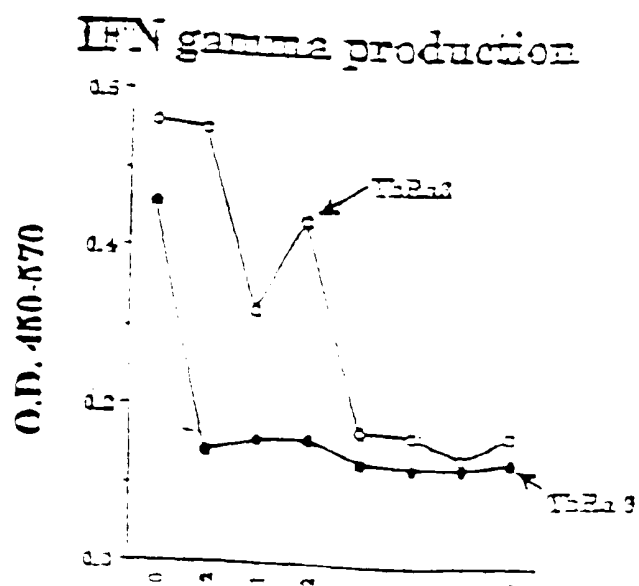
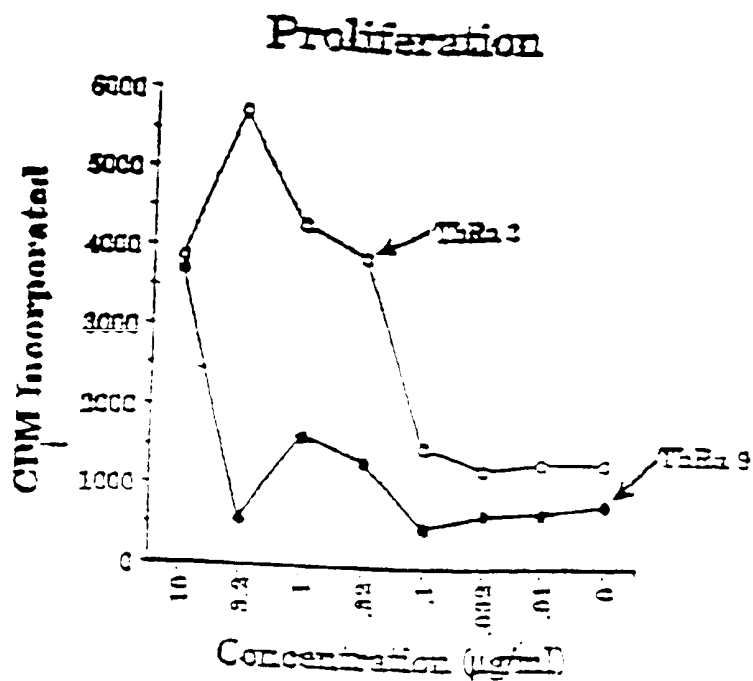
D160 T Cell Proliferation

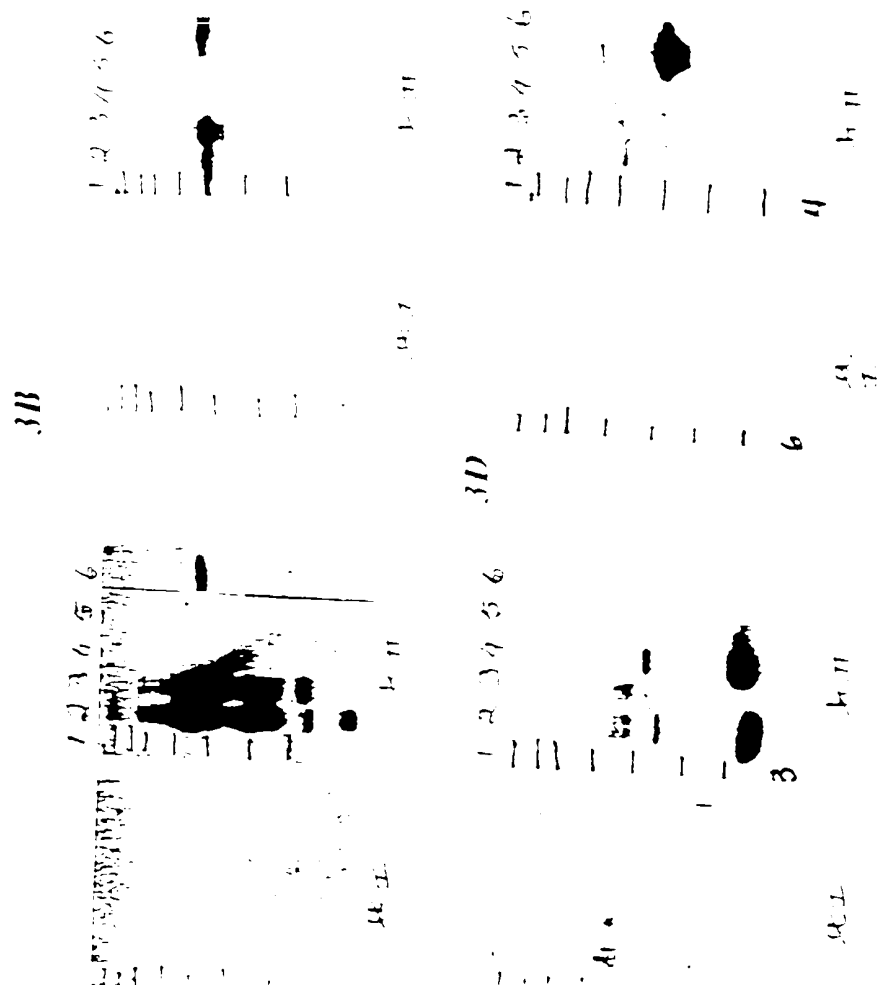


D160 IFN_g

O.D. 450-570







FIGS. 3A-D

T cell clone 131TbH9 responds poorly to CSP

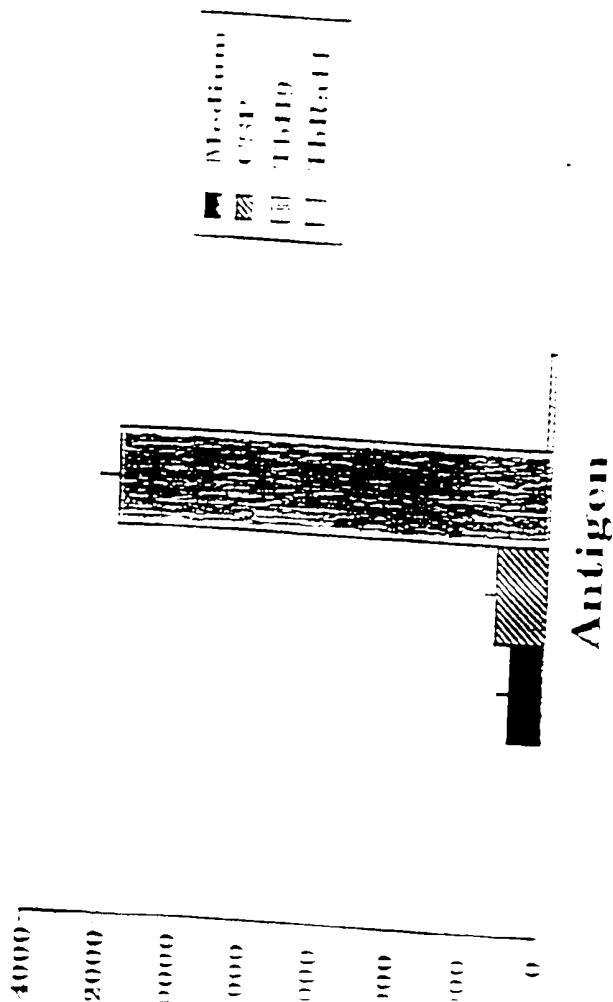


FIG. 4A

T Cell Clone PPD 800-10 IFN γ Production

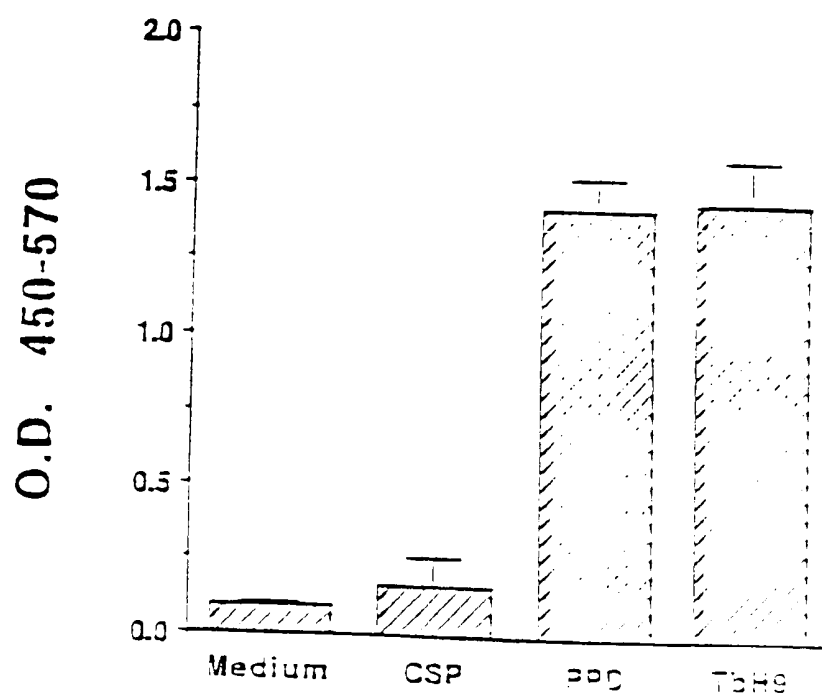
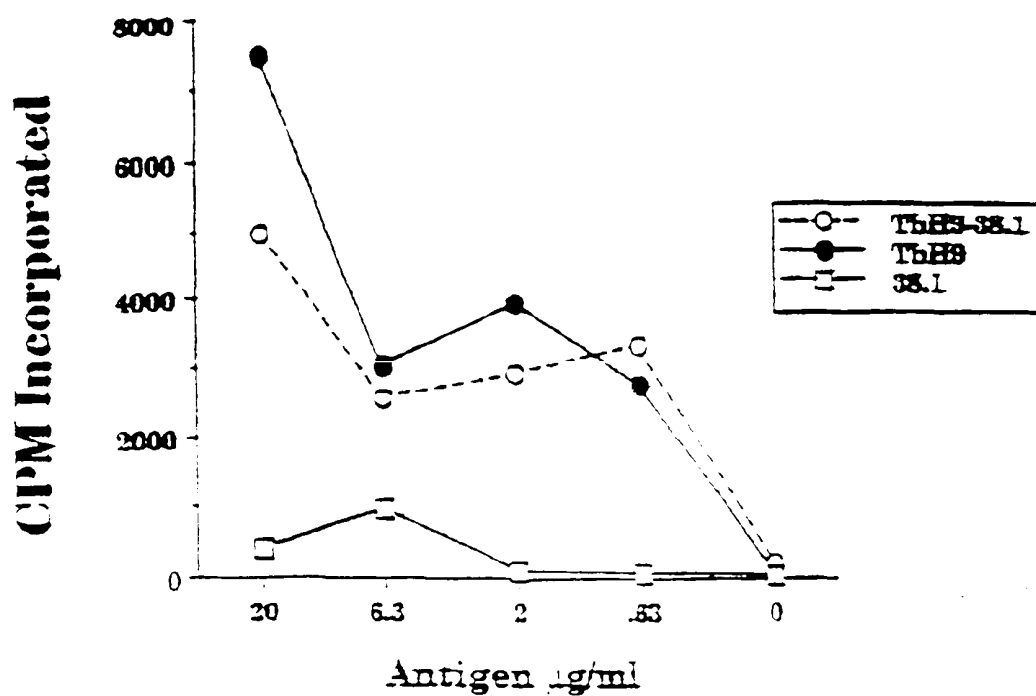
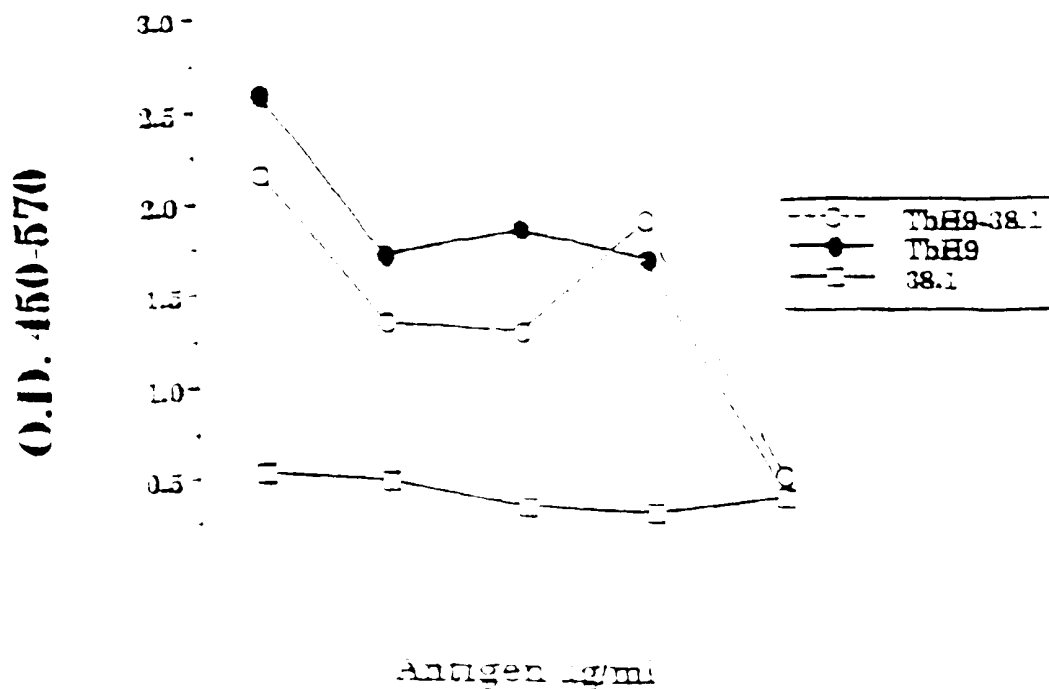


FIG. 4B

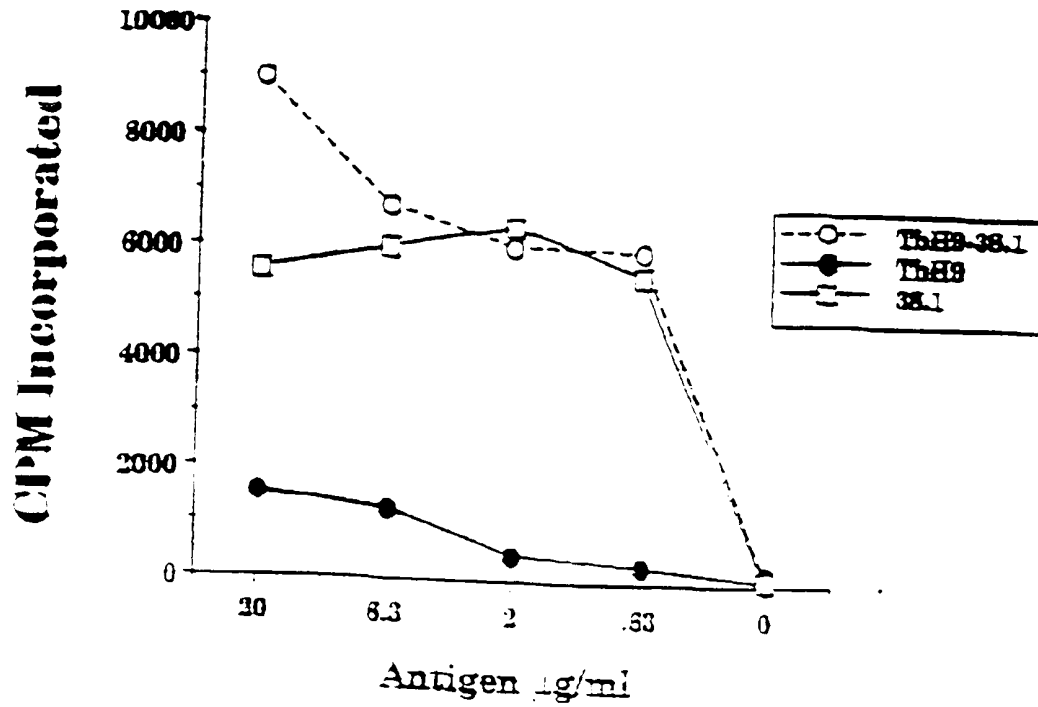
D131 T Cell Proliferation



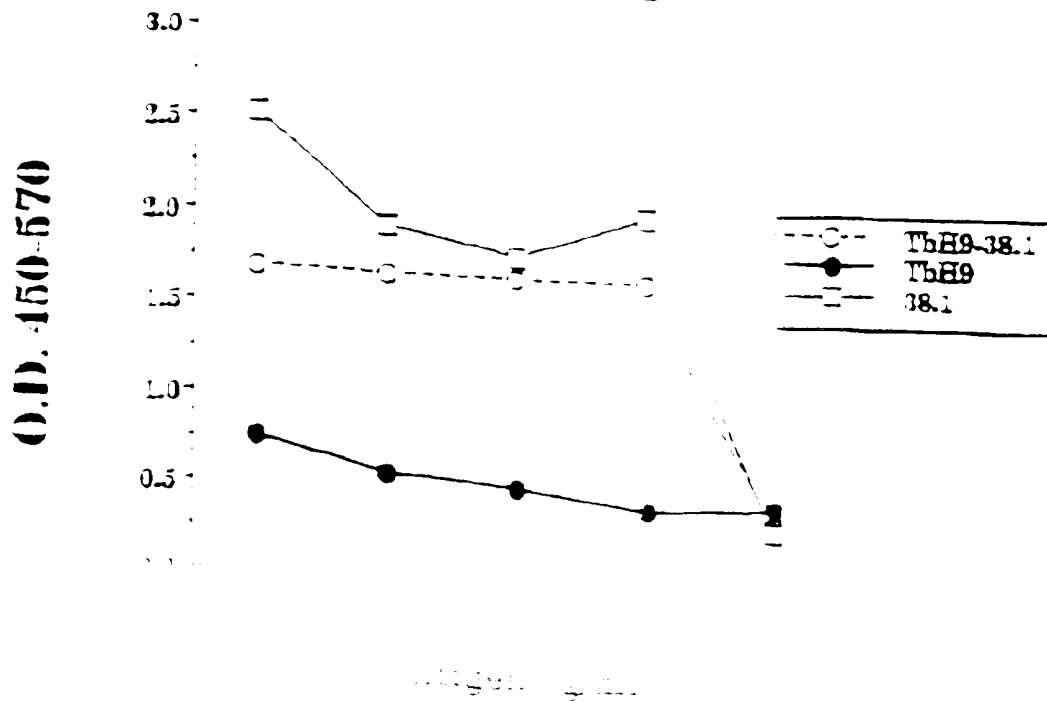
D131 IFN γ

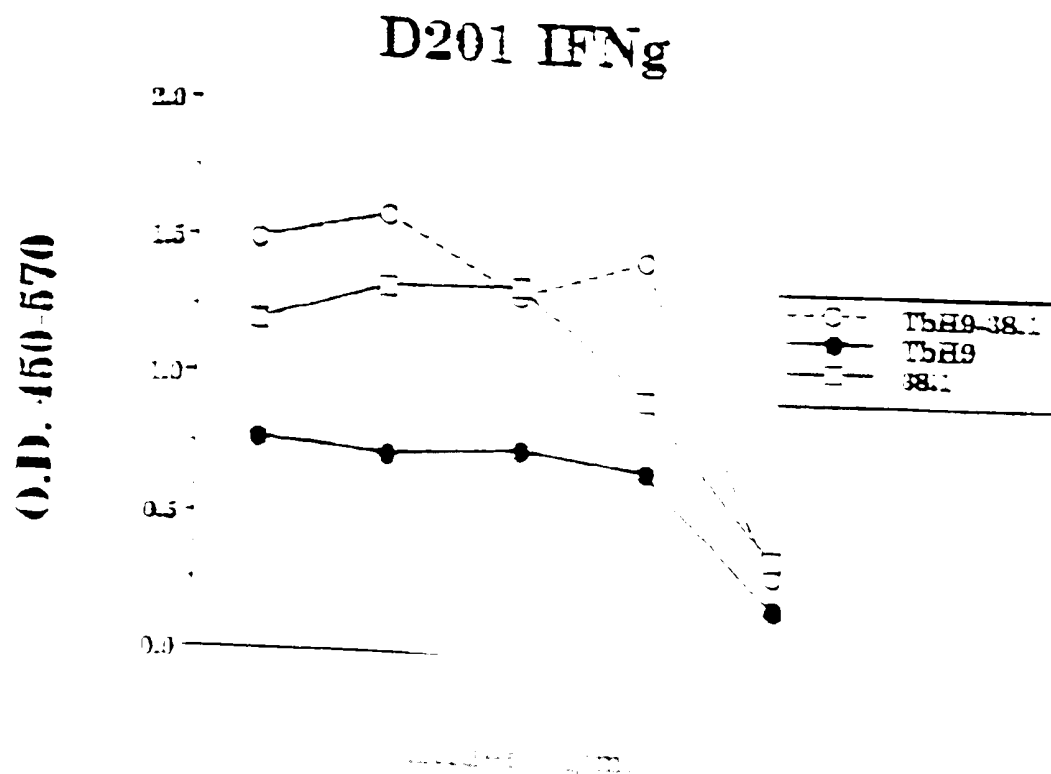
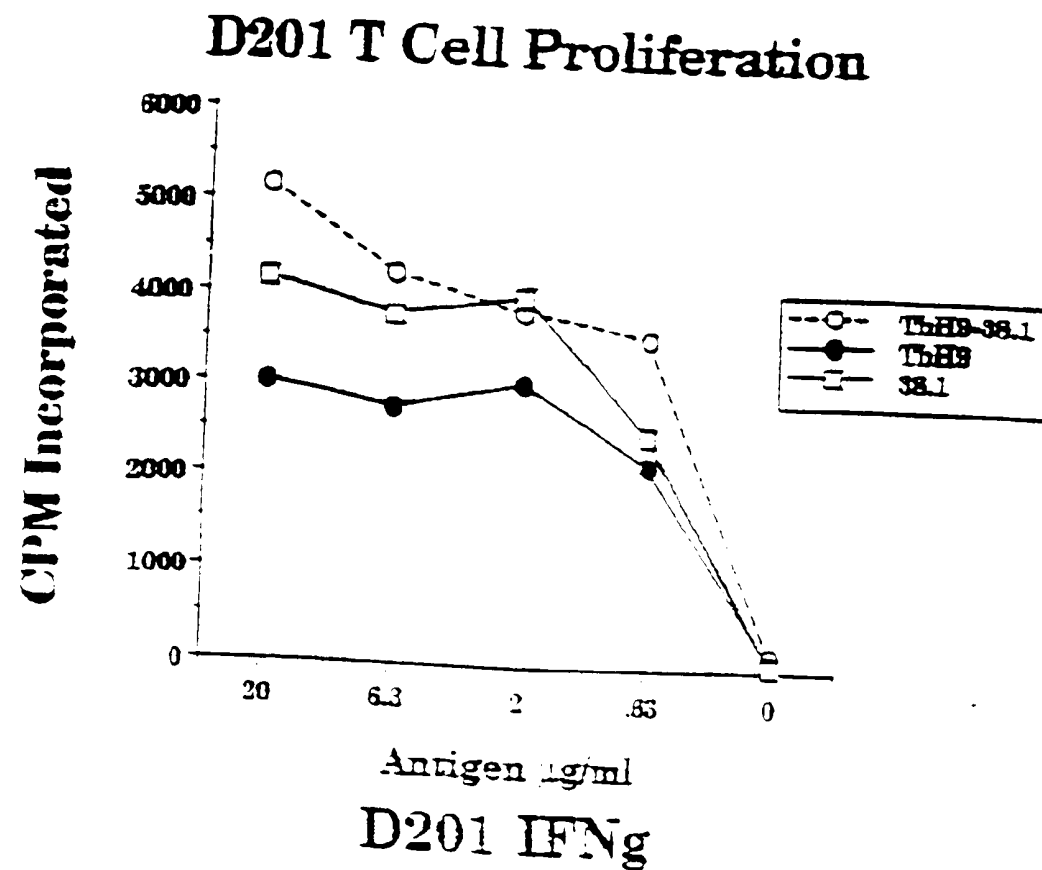


D184 T Cell Proliferation

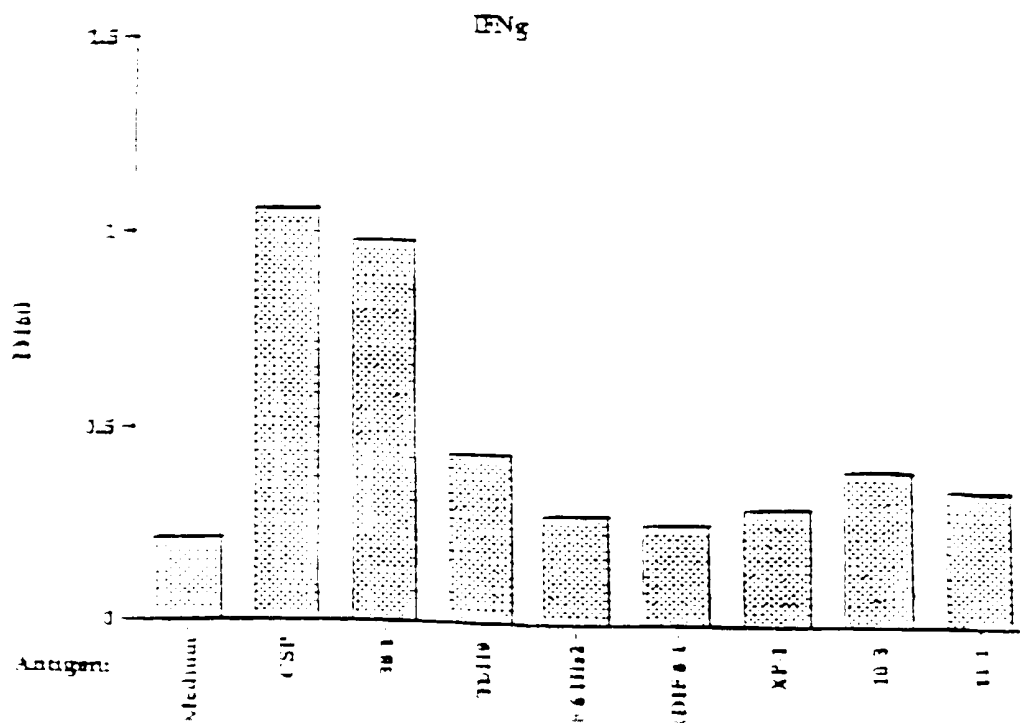
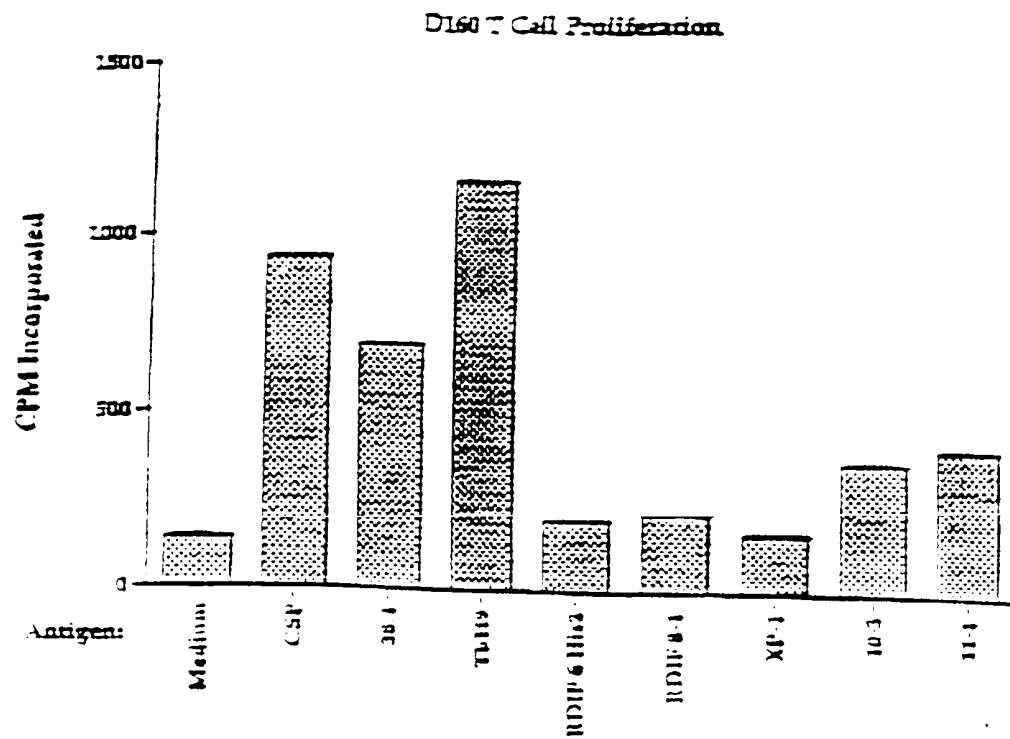


D184 IFN γ

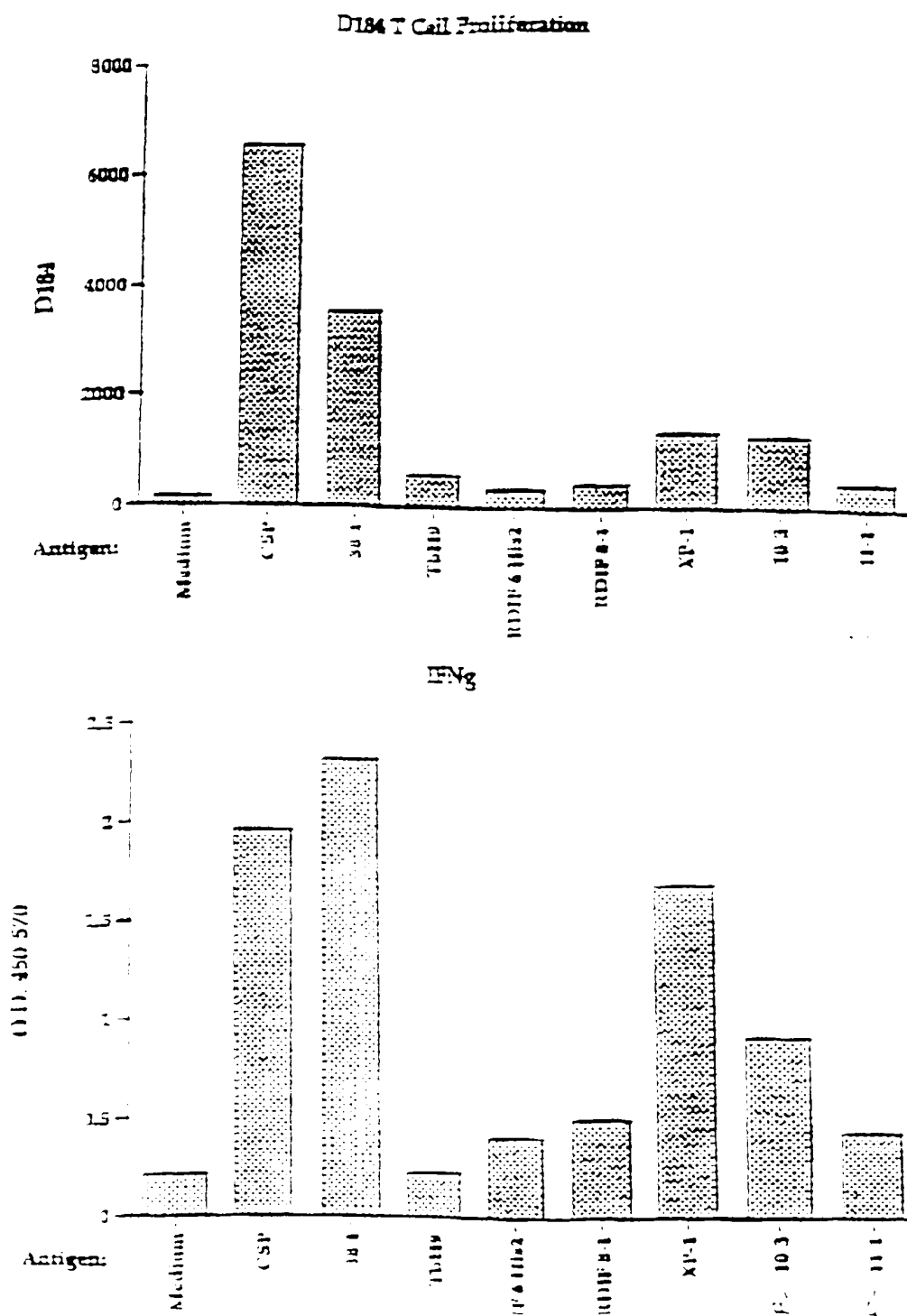




FIGS. 7-4-B

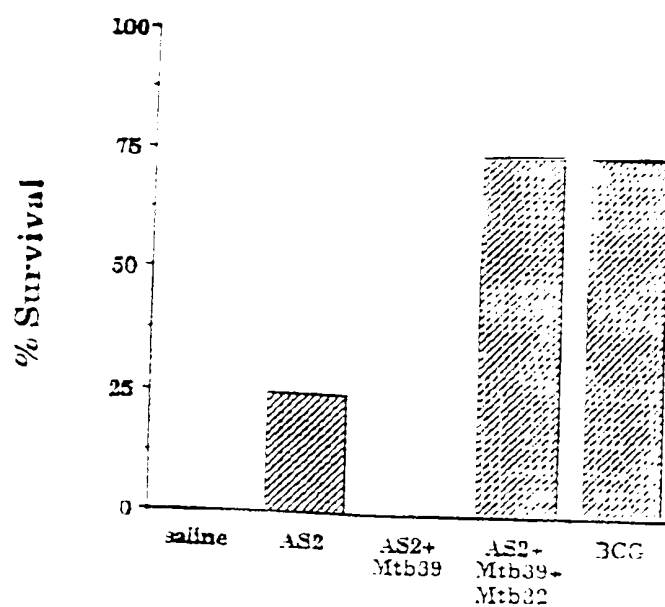


FIGS 8A-B

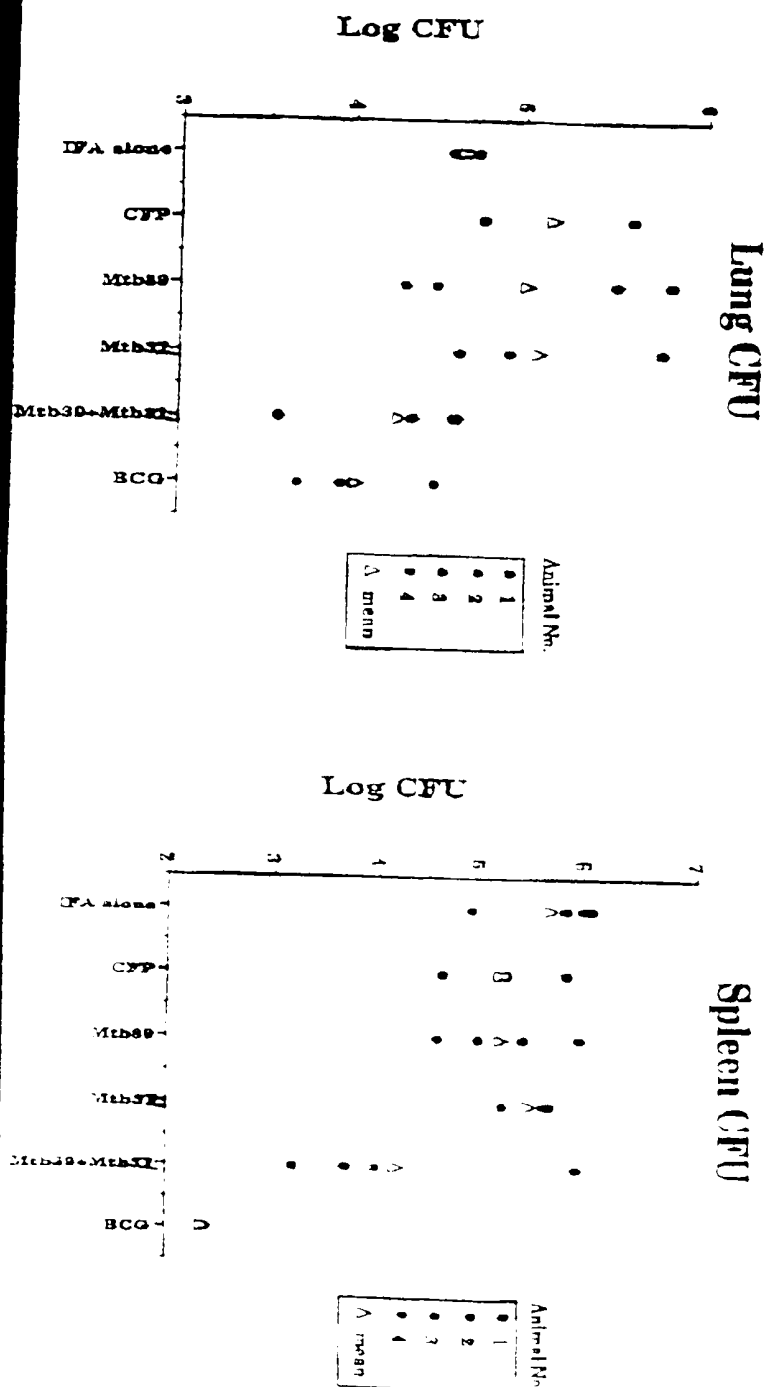


FIGS. 9A-B

**Tuberculosis: Protection of Cynomolgus Monkeys
with Recombinant Antigens of Mtb**



Aerosol TTB Challenge of Vaccinated Guinea Pigs



FIGS. 11A-B

DNA Immunized mice challenged with aerosol TB (lung CFU)

